

AD-A203 472

WHS FILE COPY

AD _____

①

**"Cloning and Expression of Genes for Dengue Virus
Type-2 Encoded-Antigens for Rapid
Diagnosis and Vaccine Development"**

ANNUAL PROGRESS REPORT

by

Radha Krishnan Padmanabhan, Ph.D.

November 17, 1987

Supported by

U.S. ARMY MEDICAL RESEARCH DEVELOPMENT COMMAND

Fort Detrick, Frederick, Maryland 21701-5012

Contract No. DAMD17-85-C-5273

University of Kansas Medical Center
Department of Biochemistry
Kansas City, KS 66103

DTIC
S E D
JAN 12 1989
H

DOD DISTRIBUTION STATEMENT

Approved for Public Release; distribution unlimited

"The view, opinions and/or findings contained in this report are those of the author and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation"

89 1 12 014

REPORT DOCUMENTATION PAGE

1. REPORT SECURITY CLASSIFICATION Unclassified			2. RESTRICTION MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION University of Kansas Medical Center		6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) Department of Biochemistry/Molecular Biology University of Kansas Medical Center 39th and Rainbow Blvd. Kansas City, KS 66103			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research and Development Command		8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAMD17-85-C-5273		
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick, Frederick, MD 21701-5012			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 63750A	PROJECT NO. 3M4-63750 D808	TASK NO. AH
			WORK UNIT ACCESSION NO. 031		
11. TITLE (Include Security Classification) Cloning and Expression of Genes for Dengue Virus Type 2 Encoded Antigens for Rapid Diagnosis and Vaccine Development					
12. PERSONAL AUTHOR(S) Radha K. Padmanabhan, Ph.D.					
13a. TYPE OF REPORT Annual		13b. TIME COVERED FROM 09/15/86 TO 09/14/87	14. DATE OF REPORT (Year, Month, Day) 1987 November 17		15. PAGE COUNT 47
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	DNA sequence analysis; Nonstructural proteins; E. coli expression; Fusion protein to beta-galactosidase; polyclonal anti-rabbit antibodies		
06	13				
06	03				
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
1.0. ABSTRACT. We carried out DNA sequence analysis of dengue virus type 2 cDNA clones #505, A4, and B2, which were cloned in our laboratory at KUMC. In addition, we sequenced the cDNA clone, pRP2 which was cloned by Dr. Robert Putnak at WRAIR. The pRP2 clone encodes the region from the carboxy terminal region of E glycoprotein to the N-terminal region of NS2A. The sequence data obtained by Maxam-Gilbert method in our laboratory was compared with the data obtained by Sanger's dideoxy sequencing method. A manuscript resulting from this collaborative work has been submitted to Virology.					
2. The complete sequence of clones #505, A4, and B2 gave rise to a total of 7406 nucleotides, so far sequenced of dengue virus type 2 genome (New Guinea strain).					
3. We chemically synthesized specific oligodeoxynucleotide primers, based on our sequence data. These primers are currently being used to complete our cDNA (continued on a separate sheet)					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Mrs. Virginia Miller			22b. TELEPHONE (Include Area Code) 301/663-7325	22c. OFFICE SYMBOL SGRD-RMI-S	

Table of Contents

1.0 ABSTRACT	1
2.0 FORWARD	2
3.0 INTRODUCTION	3
4.0 Body of the Report	4
Section I	
Characterization of the cDNA clones by Southern hybridization and DNA sequence analysis	4
1. Screening of Dengue-2 cDNA clones by Southern Hybridization.....	4
2. DNA sequence analysis of the clones pRP2, #505, A4 and B2.....	5
3. Amino Acid Sequence Deduced From DNA Sequence	5
4. Comparison of NS1 Among Various Flaviviruses.....	6
a. Rationale.....	6
b. Analysis of NS1 Sequences Among Different Flaviviruses.....	6
c. Hydropathy Plots of NS1 from various Flaviviruses	6
d. Homology of Polyprotein Segments between different Flaviviruses.....	7
Section II	
Expression of Dengue Viral Antigens In E. coli	8
1. Expression of E and NS1 epitopes	8
2. Immunoblots of fusion constructs containing E and NS1 epitopes.....	9
3. Purification of fusion proteins by affinity chromatography	9
4. DNA sequence analysis of the E-NS1 fusion constructs.....	10
5. Expression of N-terminal segment of NS3 protein in E. coli.....	10
6. Expression of C-terminal domain of NS3 in E. coli.....	11
Legend to Figures and Tables	12
Literature Cited.....	48
Personnel Supported during the Reporting Period.....	49



Accession For		
NTIS	GRA&I	<input checked="" type="checkbox"/>
DTIC TAB		<input type="checkbox"/>
Unannounced		<input type="checkbox"/>
Justification		
Distribution/		
Availability Codes		
Avail and/or		
Dist	Special	
A-1		30

1.0 **ABSTRACT** (continued from ... of DD Form 173)

3. Cloning and DNA sequence analysis.

4. We have expressed the cDNA encoding different antigens of dengue-2 virus in E. coli as fusion protein to β -galactosidase. The regions that were expressed include the C-terminal portion of E, NS1, NS3 and NS5. These fusion proteins have been purified by affinity column chromatography. The purified protein was used to immunize rabbits to produce polyclonal antisera.

5. The recombinant fusion polypeptides were analyzed by SDS-polyacrylamide for their sizes and immunoreactivities with the polyclonal mouse ascites hyper immune sera, as well as, with the monoclonal antisera against NS1 and E.

6. The fusion constructs expressing E and NS1 were verified by DNA sequence analysis to map the region which is fused with the reading frame encoding β -galactosidase.

2.0 FORWARD

The investigators have abided by the National Institutes of Health Guidelines for research involving Recombinant DNA molecules (April 82) and the Administrative Practices Supplements, as indicated in the Memorandum of Understanding and Agreement, approved by the Institutional Biosafety Committee and N.I.H.

Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

3.0 INTRODUCTION

Progress Report (September 15, 1986-September 14, 1987) can be divided into two major sections, focussing on the overall goal of 1) determining the complete sequence analysis of dengue virus genome and 2) the expression of the cloned genes, especially the surface glycoprotein E in *E. coli*.

Section I: Characterization of the cDNA clones by Southern hybridization and DNA sequence analysis were carried out. The primary sequence was compared with those of other published DNA sequences of flavivirus genomes, such as Yellow fever virus, Murray Valley Encephalitis virus, West Nile virus, Japanese Encephalitis virus. Four cDNA clones, pRP2, 505, A4 and B2 were carried out

Section II: Expression of open reading frame DNA from pRP-2 encoding C-terminal region of E and N-terminal region of NS1, pVV1 encoding the N-terminal region of NS3, #505 clone, encoding the C-terminal region of NS3 and the clone A4, encoding the NS5 protein were carried out. The methodology used for expression, purification of the polypeptides fused to β -galactosidase and immunization of the rabbit with the purified proteins for polyclonal antisera will be described under this section.

4.0 Body of the Report

Section I: Characterization of the cDNA clones by Southern hybridization and DNA sequence analysis

a. Rationale:

One of the overall objectives of the contract proposal is to sequence the entire dengue-2 virus genome. During the last ANNUAL REPORT [between September 14, 1985-September 14, 1986], the DNA sequence analysis of the regions defined by the clones pVV9, pVV1, and pVV17 was carried out (Fig. 1). The region sequenced totalled 4600 nucleotides. The clones pVV9 and pVV1 overlapped by about 450 nucleotides, whereas pVV17 and pVV1 did not. When the sequence data was compared to that of Yellow fever virus genome, it was found that the composite sequence of pVV9 and pVV1 spanned a region from the C-terminal NS1 to the N-terminus of NS3. However, pVV17 clone encoded a segment of the polyprotein from ns4b in the upstream region, all the way upto N-terminus of NS5 (Yaegashi et al., 1986). Therefore, it was necessary to fill the gap between pVV1 and pVV17, as well as get the complete sequence of NS1. The sequence analysis of additional cDNA clones and comparison of the sequence data with other flavivirus genome sequences showed that the sequenced region mapped in the C-terminus of E glycoprotein, all of NS1, ns2a, ns2b, NS3, ns4a, ns4b and upto 528 amino acids in the NS5 coding region. The total region sequenced comprised of 7446 nucleotides which is about 74% of the viral genome. This sequence data is reported here.

b. Experimental

1. Screening of Dengue-2 cDNA clones by Southern Hybridization.

During this period, many of the recombinant clones generated in our laboratory were further characterized by Southern hybridization using specific probe. The probes used were chosen from regions of DEN-2 genome which had already been sequenced such as pVV9, pVV1, and pVV17 cDNA clones. A new cDNA clone, pRP2, which was generated by Dr. Robert Putnak of Walter Reed Army Institute of Research which mapped 5' to pVV9 clone (see Fig.1) was also used as a probe for Southern hybridization to screen the clones generated in our laboratory. The results of the hybridizations are shown in Table I. We have identified among our clones those which specifically hybridized to pRP2, pVV9, pVV1, or pVV17 probes which suggested the location of these clones in the DEN-2 genome map (Fig. 1).

2. DNA sequence analysis of the clones pRP2, #505, A4 and B2.

The cDNA clone, pRP2 was generated by Dr. Robert Putnak of Walter Reed Army Institute of Research using the oligodeoxynucleotide primer synthesized in our laboratory from the sequence data of pVV9 cDNA clone. The sequence of the primer used is GTAGTCATCATCAAT which is complementary to the dengue-2 RNA in the putative coding region of ns2a (from nucleotide number of 572 to 586 in Fig.3 of Yaegashi et al., 1986). The cDNA clone pRP2 which is 2 kb long was obtained. It was then subcloned into pUC18 plasmid, so that the inserts were of amenable size for DNA sequence analysis. For example, the Bam HI fragment (500 bp), Eco RI fragments (E1 and E2, 550 bp and 800 bp, respectively), and Pst I fragment (1 Kb) were cloned at the polylinker site of pUC 18. pRP2 clone which is 2 kb long was also subjected to Bal-31 digestion to get sequential deletions from one end of the cDNA to get overlapping sequence data. This approach has been used successfully in our laboratory and the sequence of pVV9, pVV1 and pVV17 clones were carried out by this approach. Dr. Putnak sequenced this clone by using Sanger's dideoxy chain termination method and a few errors produced in each method were cross-checked and corrected. Both strands of DNA were sequenced. The strategy used for this clone is shown in Fig. 2. Similar approach was used to sequence the other three cDNA clones by Maxam-Gilbert method.

Additional sequence analysis of cDNA clones, #505, A4 and B2 were carried out. Hybridization data shown in Table I indicated that the clone #505 hybridized with both pVV1 and pVV17 clones. A4 is a ~1.7 kb cDNA clone which hybridized with pVV17 (see Table I). Similarly, B2 hybridized with A4 but not with pVV17. The order of the cDNA clone from the 5' to the 3' terminus of Dengue 2 genome is pRP2-pVV9 (pVV18)-pVV1-#505-pVV17-A4-B2. Sequence analysis at the termini of A4 and B2 confirmed this order and revealed the extent of the overlap. A4 overlapped to the extent of 1.1 kb with pVV17 and B2 overlapped with A4 about 100 bp. In addition, one terminus of B2 extends to about 220 bp 3' to A4. Sequence analysis of all these clones revealed that a total of 7448 nucleotides formed one long open reading frame from one terminus of pRP2 to that of B2 clone. The composite nucleotide sequence of all the overlapping clones is shown in Fig. 3.

3. Amino Acid Sequence Deduced From DNA Sequence

The amino acid sequence deduced from the DNA sequence is shown in Fig. 4. The composite nucleotide sequence of the clones which is 7448 bp gave one long open reading frame which spanned the entire length of 7448 bp. This result suggested that all cDNA clones, pRP2, pVV9, pVV1, #505, A4 and B2 are from the region of RNA which codes for a polyprotein precursor from which the structural and non-structural proteins are processed (Rice et al, 1985).

4. Comparison of NS1 Among Various Flaviviruses.

a. Rationale.

NS1, formerly called NV-3, is a virus-encoded, non-structural glycoprotein first described as a soluble, complement fixing antigen in dengue-infected tissue culture supernatants and in the serum of dengue-infected mice (Brandt et al., 1970). In addition to its soluble form, NS1 is also expressed on the surface of the infected cells. NS1 may play an important role in the immune response to dengue virus infection since passive or active immunization can protect mice against lethal virus challenge (Schlesinger et al.). This protection may be due to antibody-dependent, complement mediated immune cytolysis of cells which express NS1 on their surface. Understanding the structure and function of NS1 is important to develop an effective vaccine against dengue virus infection.

b. Analysis of NS1 Sequences Among Different Flaviviruses.

Comparison of the primary sequence data for NS1 from the various flaviviruses sequenced to date is carried out by the homology matrix program of Pustell and Kafatos (1984) which is commercially available from International Biotechnologies, Inc. (New Haven, Conn.). This program reveals the degree of homology and the location of homologous regions between two sequences by assigning the letters A to Z for values from 100-50% or to any minimum value arbitrarily set (50% in Fig.5). The data shows that dengue-2 NS1 bears the strongest homology with the corresponding polypeptides from dengue-1 and -4 and least homology with yellow fever virus NS1. The NS1 polypeptide from Japanese Encephalitis (JE), Murray Valley Encephalitis (MVE), West Nile virus (WN) showed intermediate homology with dengue-2 NS1. NS1 coding region also contains the well-conserved signal sequence of the type Asn-X-Ser/Thr for glycosylation which are marked by an asterisk in Fig. 3. In addition, a third potential N-linked glycosylation site is located in NS1 which is apparently not conserved. Other characteristic features of the primary structure of NS1 are as follows. There are 10 conserved cysteine residues, five of which are clustered within a 50 aa stretch from aa 280 to aa 330 relative to the N-terminus of NS1.

c. Hydropathy Plots of NS1 from various Flaviviruses

The hydropathy profiles of NS1 proteins from various flaviviruses were compared. Fig. 6 shows the result. Plots show, at each point, the running averages of the hydropathic indices of 9 consecutive amino acids. Points above the line indicate hydrophobic and points below the line, hydrophilic, residues (Kyte and Doolittle, 1982; Pustell and Kafatos, 1984). There is a striking conservation of functional domains of NS1 such as the C-terminal hydrophobic domain which perhaps functions as a membrane anchoring domain of NS1. The region of NS1 plotted in this

Fig. 6 is from amino acid residue 1-438 from the N-terminus of NS1. Fig.7 shows the hydropathy plot of the entire region sequenced. There are several unique structural characteristics of the polyprotein segment. First, the hydrophobic domain at the C-terminus of E glycoprotein, which presumably serves as the N-terminal signal sequence for NS1 (Rice et al., 1985). This domain is highly conserved among the flaviviruses. The C-terminal hydrophobic domain of NS1 is followed by ns2a and ns2b which are also extremely hydrophobic and highly conserved, although there is very little primary sequence homology. The coding regions for ns4a and ns4b are also hydrophobic.

d.Homology of Polyprotein Segments between different Flaviviruses

Complete or almost complete nucleotide sequence of two flavivirus genomes have appeared in the literature, that of Yellow Fever virus (Rice et al. 1985) and West Nile virus (Castle et al., 1986). Therefore, it was possible to compare the sequence of dengue-2 virus (New Guinea strain) we have so far obtained with these two viral genomes using alignment program of Protein Identification Resource system facility (Georgetown, Washington, DC).

DEN-2 vs WN: The amino acid sequence of the entire polyprotein segment of West Nile virus has been published (Castle et al., 1986). Alignment of the corresponding region of Den-2 sequence with that of West Nile virus polyprotein revealed the regions of the polyprotein that are highly conserved among these two flaviviruses. In addition, it allowed us to map the region of the cDNA clones which we sequenced. The alignment data is shown in Figs. 8-10. The alignment is broken into three parts for convenience. The 5' most sequence of Den-2 virus maps in the C-terminal region of E and contains 240 amino acids of this glycoprotein. Fig. 6 shows that there are many conserved regions between the two viral genomes in the NS1 region, although this homology is not perfect in the case of E, except in the C-terminal hydrophobic domain which might have a common function. The least homologous regions between different flaviviruses is the ns2a and ns2b, followed by ns4a and ns4b (Fig. 8). On the other hand, NS3 and NS5 show the highest degree of homology (Figs. 9-10).

Den-2 vs YF: Figs. 11-13 show the corresponding alignment data between Den-2 and Yellow fever virus genomes. The homology between Den-2 and YF is less striking compared to the pair of Den-2 and WN viruses. It is also evident from the homology plot by the dot matrix program in the NS1 region of the two pairs of flavivirus genomes (Fig. 5). This data show that Den-2 is closer to WN virus evolutionarily than to YF virus. The putative cleavage sites between the viral polyprotein precursors are shown as curved arrows in Figs 8-13. These sites are essentially conserved, indicating that in these flaviviruses, the protein processing occurs through a common mechanism.

% Homology between DEN-2, YF and WN viruses:

From the alignment data shown in Figs.8-13, it was possible to calculate the % homology between these three flaviviruses. The values were tabulated as the % of identical amino acids within a coding region of a viral polypeptide (see Table 2). The results are consistent with the

homology plots using dot matrix program in the NS1 region, as well as the alignment data between the different flavivirus genomes shown in Fig. 8-13.

Section II: Expression of Dengue Viral Antigens In *E. coli*

a. Rationale

One of the objectives of the Contract Proposal is to express the dengue viral antigens in *E. coli* to produce large amounts of the protein and to produce polyclonal antibodies against these antigens. To meet the goal of the proposed project on time, we initiated the expression of the different regions of the cloned cDNAs encoding different viral antigens. Since the clone pRP2 encodes the C-terminal domain of the glycoprotein, E, and all of NS1, we used this clone for the expression of these two viral antigens. We used the expression vector, pMR100 developed by Rosbach () to produce a fusion protein to *E. coli* β -galactosidase. The advantage of this system is that it is easy to screen the recombinants containing the expressing region fused to the protein coding region of β -galactosidase. The transformants are grown on an indicator plate containing McConkey agar in the presence of an inducer such as lactose, because the hybrid gene encoding a segment of dengue viral antigen and the β -galactosidase is placed under the control of lac promoter. The clones which are expressing the fusion protein are seen as red colonies. Another advantage of this system is that the fusion proteins can be purified by an affinity column in two steps to near homogeneity.

b. Experimental

1. Expression of E and NS1 epitopes

The strategy used for expression is shown in Fig. 14. The vector pMR100 tandem copies of lac promoter, a polylinker cloning site and a complete coding sequence of β -galactosidase. pMR100 is an out-of frame expression vector, so that no β -galactosidase is made from the *E. coli* strain LG90 transformed by this plasmid. pMR200 is an in-frame expression vector and active enzyme is made from the transformed cells. An insertion of an open reading frame DNA can restore the reading frame of β -galactosidase in pMR100 and produce an active enzyme in a theoretical frequency of 30%. First, pRP2 clone was digested with either PstI or HindIII and the DNA fragments were purified on a gel. It was then treated with the exonuclease Bal31 to randomize the ends so that about 30% of the DNA molecules will be able to restore the reading frame of β -galactosidase in pMR100. Transformation gave rise to several red colonies which were picked and screened by restriction enzyme cleavage, as well as by SDS-10% polyacrylamide gels to detect the synthesis of fusion protein. The region of Den-2 RNA chosen for expression (HindIII or PstI fragment in Fig. 14) contains the coding sequence for both the C-terminal E glycoprotein and

the N-terminal NS1. Therefore, if the Bal31 digestion was not extensive, the fusion construct is likely to have the coding sequences for both the antigens.

2. Immunoblots of fusion constructs containing E and NS1 epitopes

Five red colonies were analyzed by SDS-PAGE and Western blots using the polyclonal hyperimmune mouse ascites fluid raised against the dengue-2infected suckling mouse brain homogenate, polyclonal anti- β -galactosidase, monoclonal anti-E antisera (4G2) or the monoclonal anti-NS1,3E9. SDS-PAGE showed that the sizes of the fusion proteins were slightly smaller than the β -galactosidase marker, although the recombinant expression plasmid had cDNA inserts, suggesting that the fusion protein had undergone degradation to a more stable limit product in *E. coli*. The levels of these limit products were very high. In order to test whether these fusion polypeptides have the epitopes recognized by the dengue viral antigen-specific antibodies or the polyclonal anti- β -galactosidase antibodies, Western blot analyses were carried out.

Fig. 15 shows the Western blot in which the polyclonal anti- β -gal antibodies were used. Lanes 1-5 contained clones, #3, 8, 10, 16 and 3R expressing fusion proteins which had undergone degradation to a stable limit product. Lanes 6-8 show the β -gal marker protein Lane 9 contained an extract from clone 16. This result showed that the fusion protein contained the epitopes of β -gal as expected. Next, these fusion proteins were reacted with the monoclonal anti-E antibody, 4G2 (Fig.16) All the fusion constructs reacted with this antibody (lanes 3-7), although there was some non-specific reaction because β -gal also showed some reactivity (lanes 1,2,8 and 9). However, these clones did not show any reactivity towards 3H5, a different anti-E monoclonal antibody (Fig. 17). When these fusion proteins were reacted with the anti-monoclonal NS1 antibody, 3E9, (Fig. 18), only four fusion proteins, clone 8 (lane 4), clone 16 (lane 5), clone 3R (lane 6), and clone 3 (lane 7) reacted. Clone #10 did not react with this antibody. This monoclonal also showed some non-specific reactivity against β -gal protein (lanes 1,2, 3, and 9). Fig. 19 shows the immunoblot in which the polyclonal mouse hyperimmune ascites fluid was used. Only the clones 3R and 16 showed positive signal (lanes 3 and 6). It is rather surprising that the polyclonal reacted with only two clones, whereas the NS1 monoclonal reacted with four clones and E monoclonal reacted with all five fusion proteins. Further analysis with these monoclonals are necessary to confirm and explain this observation. It is possible that reactivity with these monoclonals may not be specific.

3. Purification of fusion proteins by affinity chromatography

The fusion proteins containing E and NS1 epitopes were purified by affinity chromatography using p-aminopropyl 1-thio- β -D-galactopyranoside covalently linked to agarose as the affinity matrix (Sigma Chemical Co.). The *E. coli* cell lysate containing the fusion protein was partially fractionated by ammonium sulfate precipitation (50% saturation). Subsequent to a dialysis step to

remove ammonium sulfate, the proteins were passed through an agarose affinity column as described (Germino et al., 1983). SDS-PAGE analysis of the affinity-purified fusion protein showed that it was greater than 95% pure (see Fig. 20, lane 4 and 8). Lanes 1 and 5 contained the size markers, lanes 2 and 6 contained partially fractionated fusion proteins 3R and 16, respectively before passing through the affinity column, and lanes 3 and 7 contained the pass-through fraction from the affinity column. The affinity-purified proteins 3R and 16 were used to immunize rabbits to produce polyclonal antisera against the putative E and NS1 epitopes present in 3R and 16 proteins. A high titer polyclonal antisera against the fusion protein from clone 16 has been obtained.

4. DNA sequence analysis of the E-NS1 fusion constructs

In order to determine the exact region of E and NS1 inserted into pMR100 vector to give rise to expressing clones, it was necessary to carry out DNA sequence analysis of the dengue-2 cDNA inserts after releasing them from pMR100 vector. The inserts were released by digesting with BamHI, 3'-labeled and sequenced after cutting with a second restriction enzyme. The sequence analysis was carried out by Maxam-Gilbert method. From the nucleotide sequence the amino acid sequence of the coding region for the dengue antigen was deduced. They are shown in Fig. 21. The hydrophathy plot of the region of E and NS1 present in 3R and 16 shows that there is a strong hydrophobic domain present in this region (Fig. 22), which might explain the instability of this region when expressed in *E. coli*.

4. Expression of N-terminal segment of NS3 protein in *E. coli*

Similar approach was taken to express the N-terminal region of NS3 protein of Dengue-2 virus. pVV1 clone (see Fig. 1 for the location of the cDNA with respect to the viral genome), which contained the coding sequence for the N-terminal region of NS3 polypeptide was digested with PstI to release a 1.4 kb insert. It was then digested with XhoI to give rise to a 0.9 kb and a 0.5 kb fragments which were separated by electrophoresis on agarose gel. Each fragment was individually digested with Bal31 and cloned at the SmaI site of pMR100 (see Fig. 14). Red colonies were picked and analyzed by SDS-PAGE and Western blotting technique using the polyclonal anti- β -gal or mouse hyperimmune ascites fluid antibodies. The results are shown in Figs. 23-25. Fig. 23 shows the SDS-PAGE analysis of the red colonies picked from McConkey agar plates. Lanes 1 contained the size markers, and 2 & 8 contained the control cell extract from *E. coli* LG90 transformed by pMR200 vector alone. The band shown by an arrow in this lane is the 116kDa β -galactosidase. Lanes 3 and 4 are clones generated from the 0.5 kb XhoI-PstI fragment and the lanes 5-7, and 9&10 contained the clones 3,4,5,6 and 7 respectively, generated from the 0.9 kb XhoI-PstI fragment by Bal 31 digestion. Fig. 24 shows the Western blot of the same type gel containing the samples applied in the same order. The antisera used was the polyclonal anti- β -galactosidase antibody. It can be seen that lanes 3-7 and 9 & 10 all contained

fusion protein to β -galactosidase and hence larger than the parent protein in lanes 1,2 and 8. Fig. 25 shows the Western blot of the expression clones [see lane 3 (clone PM1), lane 4 (PM2), lane 5 (PM4), lane 6 (PM5), lane 7 (PM6), and lane 8 (PM7)] all react with the polyclonal mouse hyperimmune ascites sera.

5. Expression of C-terminal domain of NS3 in *E. coli*.

Similar approach was taken to express the C-terminal domain of NS3 in *E. coli*. The cDNA clone #505 was digested with EcoRI which cleaves at nucleotide no. 3948 and 4675 of the composite sequence shown in Fig. 2 (see also for the location of EcoRI site) and the EcoRI fragment was digested with Bal 31. Another cDNA clone A4 (see Fig.1) which contains the coding sequence of NS5 was first digested with PstI to release the 1.7 kb cDNA insert, subsequently digested with NdeI to give rise to two fragments about 700 bp and a 1 kb fragments. In addition, the 1 kb fragment was digested with AccI and EcoRI to get a 400 bp fragment. The DNA fragments were purified by agarose gel and subsequently digested with Bal 31. The Bal 31 digests were then ligated to pMR100 at SmaI site. Several red colonies were screened for inserts and for fusion protein by SDS page as shown in Fig. 23. Fig.1 shows the location of the various expression clones with respect to the cDNAs #505 and A4. Fig. 26 shows the Western blot of the clones generated from this experiment. Lane 1 contained the control pMR200-transformed *E. coli* LG90 extract. Lanes 2-10 contained EBII-12 (from #505, C-terminal NS3), NBIV-5, NBIV-8, AEI-12, AEI-18, AEI-21, AEII-32, AEII-33, and AEII-34, respectively. Several fusion proteins were made in *E. coli* but they undergo degradation to limit size similar to that of β -galactosidase. Fig. 27 shows another Western blot containing additional clones screened using anti- β -gal antibody. Lane 1 again contained the control as in Fig.26, and lanes 2-10 contained the clones NBIV-7, NBIV-10, AEI-17, AEI-19, AEI-20, AEII-25, AEII-28, AEII-30, AEII-31, respectively. It is seen from these experiments that the fusion proteins expressed from this region of NS3 and NS5 were unstable in *E. coli* and underwent degradation. We wanted to investigate whether the degradation of the fusion proteins could be prevented by expressing these clones in *E. coli* JM109 in which lac promoter is tightly repressed by lacⁱ mutation causing an overproduction of lac repressor. Therefore, the expression of fusion proteins can be controlled by the addition of the inducer IPTG. Fig. 28 shows a Western blot of the expression clones, lanes 2-10. The level of expression of a selected few were compared in *E. coli* LG90 and JM109. Lanes 2-4 contained the the clone EBII-12 in LG90 (lane 2), JM109 in the absence or presence of IPTG (lanes 3 and 4). The clone AEI-18 in LG90 (lane 5) or JM109 in the absence or presence of IPTG (lanes 6 & 7) was also examined. Lanes 8 and 9 contained the control transformants obtained with pMR100 and lane 10 contained LG-90 transformed by the clone AEII-27. The results show that transfer of plasmids to JM109 had no significant effect on the degradation, suggesting that this regions might be inherently unstable in *E. coli*. The expression clone EBII-12 (Fig.26, lane 2), and AEII-32 (Fig. 26, lane 8) were chosen for further purification and characterization. Western blot using the polyclonal mouse hyperimmune ascites fluid showed that these two clones were immunoreactive (data not shown).

Legend to Figures

Fig. 1: The map of the region sequenced indicating the location of the various clones either isolated in our laboratory or at WRAIR (pRP2). The terminus of pRP2 at the left extreme includes 240 amino acid coding region of E glycoprotein. The clone B2 maps about 526 amino acids into NS5.

Fig. 2: The strategy for sequencing pRP2 clone. The dotted arrows indicate dideoxy sequencing method used for sequencing and solid arrows indicate Maxam-Gilbert method. The arrowhead indicates the direction of sequencing and the length of the arrow indicates that of the region sequenced. The letters over the horizontal lines denote the restriction enzyme cleavage sites on the cDNA coding for NS1. B=BamHI; E= EcoRI; P= PstI; H= HindIII;

Fig. 3: The composite sequence of several clones sequenced. The total length is 7448 nucleotides.

Fig. 4: The amino acid sequence deduced from the nucleotide sequence. There is one long open reading frame for the entire length of the region sequenced.

Fig.5: Homology of NS1 region among various flaviviruses. The dot matrix program of Pustell and Kafatos (1984) was used. The amino acid sequences of NS1 from Yellow fever (Rice et al., 1985), Dengue 1 (Dr. Peter Mason), Dengue 4 (Dr. Lai), Japanese Encephalitis (McAda et al., 1987), Murray Valley encephalitis, MVE (Dalgarno et al., 1986, West Nile fever (Castle et al., 1986), The minimum homology value for these plots is arbitrarily set at 50%. The degree of homology is indicated by letters A-U. The assigned range in this case is 3. The amino acid residues are evaluated in overlapping groups whose size is set by the variable (range), which is the number of residues on either side of the central residue included in the group.

Fig.6: Hydropathy profiles of the NS1 proteins. Plots show at each point, the running averages of the hydropathic indices of 9 consecutive amino acids. Points above the line indicate hydrophobic and points below the line, hydrophilic residues (Kyte and Doolittle, 1982; Pustell and Kafatos, 1984). YF, yellow fever; Den-1-4, dengue-1-4; JE: Japanese Encephalitis; MVE: Murray Valley Encephalitis; WN: West Nile fever virus. The region of NS1 plotted in this diagram was from amino acid residue 1-438.

Fig. 7: Hydropathy profile of the total amino acid sequence shown in Fig. 4.

The parameters chosen were the same as described in Fig. 6. The total amino acid sequence of deduced from all the cDNA clones so far sequenced is shown in Fig.4.

Fig. 8-10: Alignment of Den-2 sequence with WNV in the corresponding region. The alignment program of Protein Identification Resource System Facility (Georgetown, Washington DC.) was used.

Fig. 11-13: Alignment of Den-2 with YF. Same as in Fig. 8-10, except that Den-2 and YF were compared.

Fig. 14: Strategy for expression of Dengue-2 open reading frame DNA.

The pMR100 vector is an out of frame vector which makes no detectable level of β -galactosidase in *E. coli* transformed by this plasmid. However, when an open reading frame DNA whose ends are randomized by Bal 31 digestion is inserted into the polylinker site in the vector, the frequency of restoring the reading frame of the vector theoretically becomes 30%. The expressing clones are selected by plating the transformants onto a McConkey agar plate.

Fig. 15: Western blot of the expressing clones.

The clones were grown and the bacteria were lysed and loaded on a 10% polyacrylamide gel and electrophoresed. The protein bands were transferred to a nitrocellulose filter paper and blocked with milk and reacted with first the primary antibody (anti- β -gal in this case), followed by a second antibody conjugated to alkaline phosphatase. The latter was detected using a chromogenic substrate.

Fig. 16: Western blot using monoclonal anti-E antibody (4G2).

The conditions were the same as in Fig. 15 except the primary antibody is different. Lane 1=pMR200; lane 2= pMR100; lanes 3-7, clones 3R, 3, 16, 8, 10; lane 8=pMR200; lane 9=purified β -gal; lane 10= size markers.

Fig. 17: Western blot using monoclonal anti-E antibody (3H5).

The conditions were same as in Fig. 16, except the monoclonal antibody was 3H5 which is specific to a different epitope of E glycoprotein.

Fig. 18: Western blot using anti-NS1 monoclonal antibody (3E9).

The conditions were same, except that the anti-NS1 monoclonal was used. lanes 1-9 contained, pMR200, purified β -gal from pMR200-transformed cells, *E. coli* extract transformed by pMR100, clone 8, clone 16, clone 3R, clone 3, clone 10 and pMR200 control.

Fig. 19: Western blot using polyclonal mouse hyperimmune ascites fluid

The lanes 1-8 contained, β -gal, clone 16, clone 10, pMR100, clone 8, clone 3R, clone 3, pMR200.

Fig. 20: Purification of fusion proteins from 16 and 3R using affinity column chromatography.

The affinity column matrix was p-aminopropyl-1-thio- β -D galactopyranoside (Sigma Chemical Co.). The *E. coli* cell extract was prepared and partially purified by ammonium sulfate precipitation. It was then passed through an affinity column and eluted from the column using a pH 10 buffer. Lanes 1 and 5= size markers, lanes 2 & 6, ammonium sulfate fraction; lanes 3 & 7 pass through fraction; lanes 4 & 8 = the column purified material; lane 4=clone 16; lane =clone 3R.

Fig. 21: Deduced amino sequence of fusion protein from E and NS1.

The cDNA insert into pMR100 to give rise to a fusion construct which express a limit product in *E. coli* was characterized by DNA sequence analysis to determine the length and amino acid sequence of the portion of dengue-2 E and NS1 antigenic domains.

Fig. 22: Hydropathy profile of the E and NS1 fusion polypeptides

The two expressing clones 3R and 16 which presumably antigenic region of E and NS1 fused to β -gal was analyzed for their hydrophobicities. There is a prominent hydrophobic region in each expressing clone which might be contributing to the instability of the fusion proteins which got degraded to a limit size.

Fig. 23. SDS-PAGE analysis of NS3 Expression clones.

The NS3 region was expressed in pMR100 vector as described in the text. The clones were analyzed by SDS-PAGE. Lanes 1-10 contained protein size markers, pMR200 extract, clones PM1-NS3, PM2-NS3, PM3-NS3, PM4-NS3, PM5-NS3, pMR200 extract, PM6-NS3, and PM7-NS3, respectively.

Fig. 24. Western blot of NS3 expression clones using polyclonal anti- β -gal antibody

Western blots were carried out as described in Fig. 15. Lanes 1-10 contained, protein size markers, pMR200 extract, PM1-NS3, PM2-NS3, PM3-NS3, PM4-NS3, PM5-NS3, pMR200 extract, PM6-NS3 and PM7-NS3, respectively. The fusion protein migrates slower than the β -gal band in lanes 1,2, and 8.

Fig. 25. Western blot of NS3 expression clones using polyclonal mouse hyperimmune ascites fluid.

Western blot was carried as described before. Lanes 1-9 contained pre-stained protein size markers, pMR200 extract, PM1-NS3, PM2-NS3, PM4-NS3, PM5-NS3, PM6-NS3, PM7-NS3, and β -gal (purified from pMR200 transformed *E. coli*).

Fig. 26. Western blot of NS3 and NS5 expression clones.

Western blot was carried out as described in Fig. 15. The antibody used was anti- β -gal antibody. The lanes 1-10 contained the extracts from fusion constructs, AEII-34, AEII-33, AEII-32, AEI-21, AEI-18, AEI-12, NBIV-8, NBIV-5, EBII-12 (NS3 region), and β -gal control, respectively. NB and AE refer to the fragments generated from NS5 coding region by NdeI-PstI (1 kb fragment) or AccI-EcoRI double digestion of A4 cDNA clone of Den-2.

Fig. 27. Western blot of NS3 and NS5 expression clones.

Western blot was carried out as described in Fig. 26. The lanes 1-10 contained the extracts from the fusion constructs (all from NS5 coding region), AEII-31, AEII-30, AEII-28, AEII-25, AEI-20, AEI-19, AEI-17, NBIV-10, NBIV-7, and β -gal from pMR200 extract, respectively.

Fig. 28. Western blot of NS3 and NS5 expression clones in LG90 or JM109.

Anti- β -gal antibody was used. The expression plasmids from NS3 and NS5 regions were transferred from the original strain of LG90 to JM109 in order to compare the stability of the fusion proteins expressed from these two strains by Western blot. The lanes 1-10 contained extracts prepared from *E. coli* transformed by the clones, AEII-27 (LG90), pMR100 extract from JM109 transformed cells (-IPTG, white colony), pMR100 extract from JM109 transformed cells (-IPTG, red colony, possibly a revertant), AEI-18 (JM109, +IPTG), AEI-18 (JM109, -IPTG), AEI-18 (LG90), EBII-12 (JM109,+IPTG), EBII-12 (JM109, -IPTG), EBII-12 (LG90), and β -gal control, respectively.

Legend to Tables

Table1. Southern hybridization of cDNA clones using specific probes.

Den2 cDNA clones isolated in our laboratory were screened using the well-characterized cDNA clones, pRP2, pVV9, pVV1 and pVV17. A + sign indicates a positive signal. From this result, clone #505, A4 and B2 were picked for sequence analysis.

Table2. Homology between Den2:YF and DEN2:WNV

The values are expressed as % calculated from the number of identical amino acids and the total number of amino acids present in a particular coding region. The coding regions are indicated by bent arrows in Figs. 8-10 for the various structural and non-structural proteins. The partial coding region for E was not taken for the homology index; however, the partial coding region for NS5 was taken because we have sequenced a portion of it.

Fig. 1

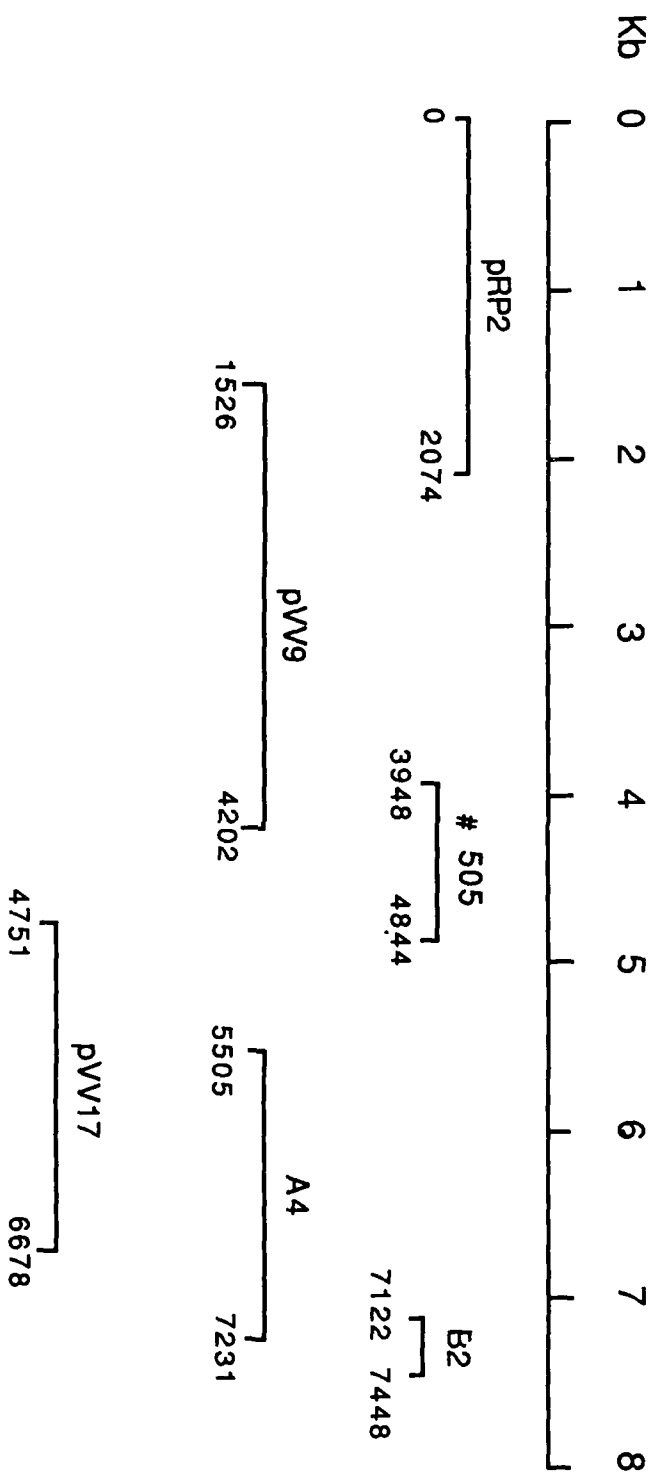
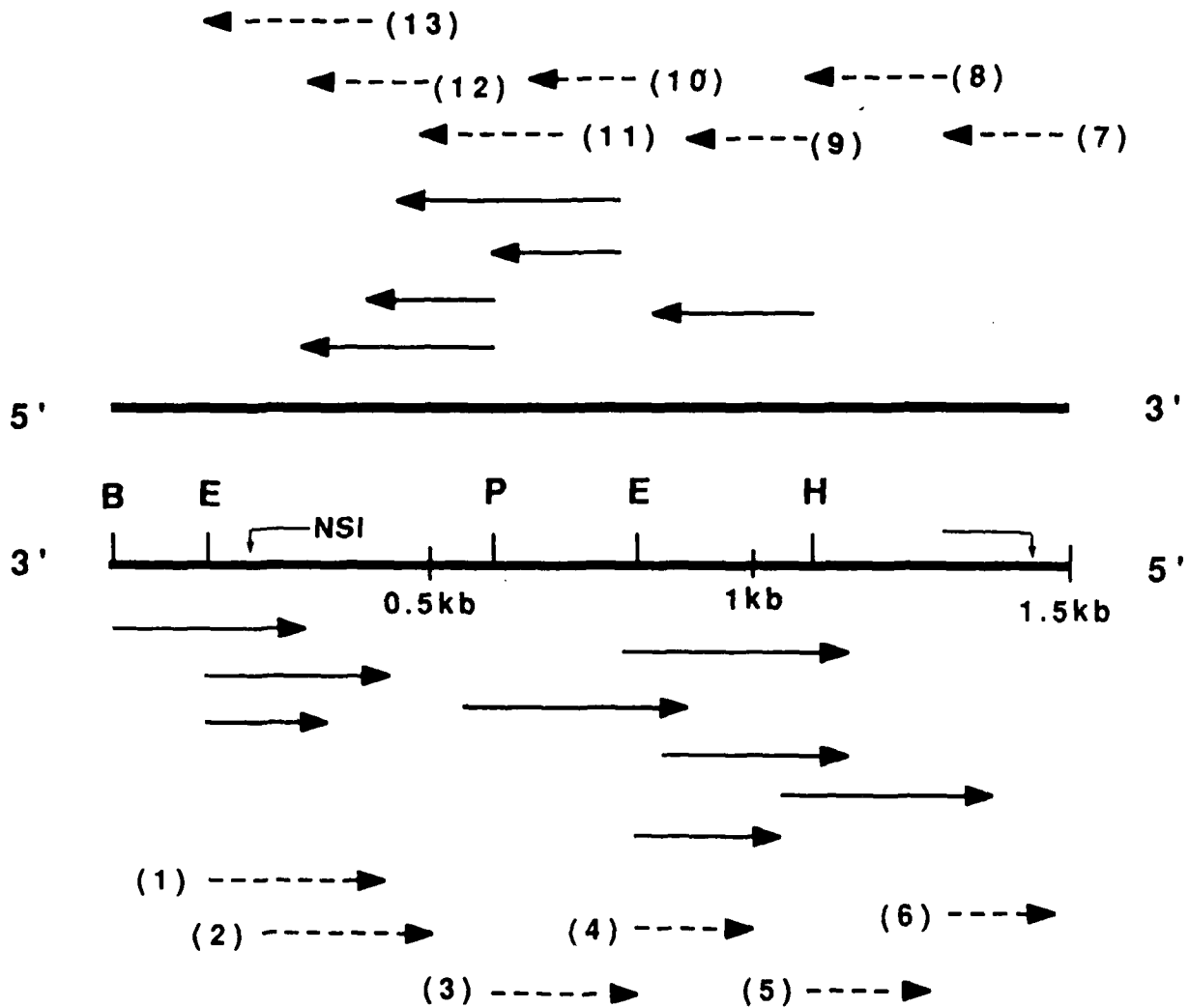


Fig.2



1 AAGGAGGGG GATGACAC AGCACTACA GGGGACACG AATTCACAT CTATAGAGG
61 AACTTACTGT TACAGGACA TCTCAAGTGT AGGCTGAGGA TGGACAAACT ACAGCTCAAA
121 GGAACTGTAT ACTTATATCT CACAGGAAAG TTTAAAGTTG TGAAGBAAT ACAGAAAACA
181 CAACATGGAA CAATAGTTAT CAGAGTACAA TATGAAAGGG ACGTCTCTCC ATCTAAGATC
241 COTTTTGAGA TAATGGATTT GGA AAAAAGA CATCTTTTAT GTGGCTGAT TACAGTCAAT
301 CCAATCGTAA CAGAAAAAGA TAGCCAGT C AACATAGAAG CAGAACTCTC ATTCCGAGAC
361 AGGTACATCA TCATAGGAGT AGAGCCGGGA CAATTGAAAG TCAACTGCTT TAAGAAAAGGA
421 AGTTCTATCG GCAAAATGAT TGAGACAACA ATGAGGGGAG CGAAGAGAAT GGCATTTTA
481 GGTGACACAG CTGGGATTT TGGATCTGG GAGGAGTGT CTACATCTAT AGSAAAGGCT
541 CTCCACCAAG TTTTGGGAGC AATCTATGG GDTGCTTBA GTGGGCTCTC ATGGATTATG
601 AAAATGCTCA TAGGAGTCAT TATCAGATGG ATAGBAATGA ATTCAGGCAG CACTCACTG
661 TCTGTGTCACT AGTATTGGT GGGAGTCTGT ACGCTGTATT TGGGACTTAC GGTCCAGGCT
721 GATAGCTGGT CCGTGTGGAG CTGGAAAAAC AAAGAACTGA AGTGTGGGAG TGGGATTTTC
781 ATCACAGACA ACGTGCACAG ATGGACAGAA CAATACAAGT TCCAACCAGA ATCCCTTCA
841 AAACAGCTT CAGCTATCCA GAAAGCTCAT GAAGAGGGCA TTTGTGGAAT CCGCTCAGTA
901 ACAAGACTGG AAAATCTGAT GTGGAAACAA ATAACACCAG AATGAAATCA CATTCTATCA
961 GAAAAAGAGG TGAAGTTGAC TATTATGACA GGAGACATCA AAGBAATCAT GCAGGCAGGA
1021 AAACGATCTC TGCAGCCCCA GCCCAGTGA CTGAAGTATT CATGAAAAA ATGGGCAAAA
1081 GCGAAAAATGC TCTCTACAGA GTCTCATAAC CAGACCTTTC TCATGAAAG CCCGAAACA
1141 GCAGAACTCC CCAACACAAA CAGAGCTTGG AATTCCGCTG AAGTTGAAAG CTATGGCTTC
1201 GGAGTATTCA CCACCAATAT ATGGCTAAAG TTGAGAGAAA AGCAGGATGT ATCTGCGAC
1261 TCAAAAATCA TGTCAGCGGC CATAAAAGAC AACAGAGGCT TCAAGGCA TAATGGTAA
1321 TGGATAGAAA GTGCACTCAA TGACACATGG AAGATAGAGA AAJCTCTTT CATCGAATC
1381 AAAAGCTGCC ACTGGCCAAA GTCACACACC CTCTGGAGTA ATCTAGTCTT AGAAAATGAG
1441 ATGATAAATC CAAAGAATTT CGCTGGACCA GTGTCAACA ACBAATAGG ABAAGTAA
1501 CACACACAAA CAGCAGGACC ATGGCATCTA GGTAGCTTC ABAATGAAAT TBAATTCAG
1561 GAAGCAACCA CAGTGGTGT GACTCAGGAC TGTGGAATA CAGGACTCTC TTAAGAA
1621 ACTACTGCTT CTGGAAAACT CATAACAGAA TGGTGTGGC GATCTTCCAC ATCTCAAA
1681 CTAAGATACA GAGGTGAGGA CCGATCTGG TACGGGATGG AAATCAGAGC ATFGAAGAG
1741 AAAGAAGA ATTFGCTCAA CTCTTGGTC ACAGCCGAC ATGGGCAGAT TGAACCTCT
1801 TCACTAGGAG TCTTGGGAAT GGCATTGTT CTGGAAGAAA TGCTCAGGAC CCGAGTACA
1861 ACGAAACATG CAATACTACT AATTGCATTT TCTTTTGA CATTGATCAC AGGCAACATG
1921 TCTTTTAGAG ACGTGGGAG AGTGATGGTT ATGTTGGGCT TACTATCAC GATGACATA
1981 GGTATGGGGT TGACTTATCT TCCCTACTA GCAGCCTCA AACTCAGACC AACTTTTGA
2041 GCTGGACTAC TCTTGAGAAA GTTGACCTCC AAGGAAITGA TGATGACTAC CAYAGGAATC
2101 GTACTCCTCT CCCAGAGCAC CATACCAGAG ACCATTCTTC AACTGACTSA TGCTTACCC
2161 TTGGCCATGA TGGTCTTAA AATGGTGAGA AAAATGGAAA AGTATCAATT CCJATGACT
2221 ATCATGGCTA TCTTGTGGT CCAAAATGCA GTGATATTAC AAAACCCATG GAAAGTGAT
2281 TGCACAATAT TGGCAGTGGT GTCCGTTTC CCACTGTTCT TAACATCTC ACACACAAA
2341 GCGGATGGA TACCATTAGC ATTGACGATA AAGGGTCTA ATCCAACAGC TATTTTCTA
2401 ACAACCTTT CAAAGAACCA CCAAGAAAAG AGCTGGCCAC TAAATGAGGC TATCATGGA
2461 GTCGGGATGG TGAGCATTIT GGCCAGTTCA CTCTAAAGA ATGACATTC CCGGACAAA
2521 CCATPAGTGG CTGGAGGGCT CTTCACTCTG TGCTACGTGC TCACTGGAGC ATGGAGGAT
2581 TTGGAACTGG AGAGACCGC CGATGTCAA TGGGAAGATC AGGCAGAGAT ATAGNATCC
2641 AGTCCAATCC TGTCAATAAC AATATCAGAA GATGUTAGCA TGTGATAAAA AAACGAACAG
2701 GAAGAACAAA CACTGACCAT ACTCATCAGA ACAGGATTGC TGGTGATCTC AGGACTTTT
2761 CCTGTATCAA TACCAATCAC GGCAGCAGCA TGGTACCTCT GGGAAATGAA CAAACAAAG
2821 GCTGGAGTAT TGTGGGATGT CCTTCACCC CCACCCGTTG GAAAGGCTA ACTGGAAGAT
2881 GGAGCCTATA GAATCAAGCA AAAAGGGATT CTTGGATATT CCCAGATCG AGCCGGAGTT
2941 TACAAAGAAG GAACATCCA TACAATGTGG CATGTCAAC GCGGCGTGT TCTAATCCAT
3001 AAAGGAAAAG GGATTGAACC ATCATGGGCG GACGTTAAGA AAGACCTEGT GTGCTGTGAA
3061 GGAGGCTGGA AGCTAGAAGG AGAATGGAAG GAAGGAGAAG AAGTCCAGGT CTTEGCATTE
3121 JAGCCTGGAA AAAATCCAAG AGCCGTCCAA ACAAAACCTG GTCTTTTCAA AACCAACCC
3181 GGAACCATAG GTGCCGTATC TCTGGACTTT TCTCCTGGAA CCTCAGGATC GCAATCATC
3241 GACAAAAAAG GAAAAGTTGT GGTCTTTAT GGTATGGTG AGTAATGAG ACAATCAGA GATCGAAGAT
3301 TATGTGAGTG CTATAGCCCA AGCTGAAAAA ATTGACCATC ATGGACCTCC ACCCAGGAGC GGGAAAGAGC
3361 GACATTTTTTCAAAGAGAAA ATTGACCATC ATGGACCTCC ACCCAGGAGC GGGAAAGAGC
3421 AAGAGATACC TTCCGGCCAT AGTCAGAGGG GCTATAAAA GGGGCTGAG GACATTAATC
3481 CTGGCCCCCA CTAGAGTCTT GGCAGCTGAA ATGGAGGAAG CCTAAJAGG ACTTCAATA
3541 AGATAACAAA CCCCAGCCAT CAGAGCTGAG CACACCGGGC GGGAGATCT GAGTGCCAAA TACAACCTG
3601 TGTCAATGCA CATTCACTAT GAGGCTGCTA TCACCAGTTA GAGTGCCAAA TACAACCTG
3661 ATCATCATGG ACGAAGCCCA TTTACAGAC CCAGCAAGTA TAGCGGCTAG AGGATACATC
3721 TCAACTCGAG TAGAGATGGG TGAGGCAGCT GGGATTTTCA TGACAGCCAC TCTCCGGGA
3781 AGCAGAGACC CATTCCCTCA GAGCAATGCA CCAATCATGG ATGAAGAAAG AGAAATCCCT
3841 GAACGTTCTT GGAGTTCTGG ACATGAGTGG GTCACGJATT TTAAGGGAA BACTGTTTGG
3901 TCTGTTCCAA GTATAAAAGC AGGAAATGAC ACAGCAGCTT GCTGAGAAA AAATGGAAG
3961 AAATGATAC AACTCAGTAA BAAGACCTTT GATTCTGAT ATCTAGGAA TCACTGAA

3721	TCAACTCGAG	TAGAGATGGG	TGAGGCAGCT	GGGATTTTCA	TGACAGCCAC	TCCCTCCSHGA
3781	AGCAGAGACC	CATTCCCTCA	GAGCAATGCA	CCAATCATGG	ATGAAGAAAG	AGAAATUCCT
3841	GAACGTTTCGT	GGAGTTCTGG	ACATGAGTGG	GTCACGGATT	TAAAGGGAA	CACTGTTTGG
3901	TTCGTTCCAA	GTATAAAAGC	AGGAAATGAC	ACAGCAGCTT	GCCGTGAGAAA	AAATGGAAAG
3961	AAAGTGATAC	AACTCAGTAG	GAAGACCTTT	GATTCTGAGT	ATGTCAAGAC	TAGAACCAAT
4021	GATTGGGAC	TCGTGGTCAC	AACTGACATT	TCAGAAATGG	GTGCCAACTT	CAAGGCTGAG
4081	AGGTTTATAG	ACCCAGACG	CTGCATGAAA	CCAGTTATAC	TAACAGATGG	TGAAGAGCGG
4141	GTGATCCTGG	CAGGACCTAT	CCAGCTGACC	CACTCTAGTG	CAGCACAAAG	AAGAGCGAGA
4201	ATAGGAAGAA	ATCCAAAAAA	TGAAAATGAC	CAGTACATAT	ACATGGGGGA	ACCTCTGGAA
4261	AATGATGAAG	ACTGTGCACA	CTGGAAAGAA	GCTAAAATGC	TCCCTAGATAA	CATCAACACA
4321	CCTGAAGGAA	TCATTCTAG	CATGTTCGAA	CCAGAGCGTG	AAAAGGTGGA	TGCCATTGAT
4381	GGTGAATACC	GCTTGAGAGG	AGAAGCAAGG	AAAACCTTTG	TGGACCTAAT	GAGAAGAGCA
4441	GACCTACCAG	TCTGGTTGGC	CTACAGAGTG	GCAGCCGAAG	GCATCAACTA	CGCAGACASA
4501	AGGTGGTGTT	TTGATGGAAT	TAAGAACAAC	CAAATCTTGG	AAGAAAATGT	GGAGGTGGAA
4561	ATCTGGACAA	AAGAAGGGGA	AAGGAAGAAA	TTAAAACCCA	GATGGTTGGA	TCCCATGATC
4621	TACTCTGACC	CACTGGCGCT	AAAGGAATTC	AAGGAGTTTG	CAGCTGGAAG	AAAGTCCCTG
4681	ACCCTGAACC	TAATCACAGA	AATGGGTAGG	CTTCCAACTT	TCATGACTCA	GAAGGCAAGA
4741	GACGCCTGG	ACAACCTTAG	AGTGCTGCAC	ACGGCTGAAG	CAGGAGGAAG	GGGTACAAAT
4801	CATGCTCTCA	GTGAACCTGC	GGAGACCTTG	GAGCAATTGC	TTTTACTGAT	ACTTCTGGCT
4861	ACAGTCACAG	GAGGAATCTT	TTTATTCTTG	ATGAGCGGAA	GGGGTATAGG	GAAGATGACT
4921	CTGGGAATGT	GCTGCATAAT	CACGGCTACT	ATTCTCCTAT	GCTACGCACA	AATAAGCCA
4981	CACTGGATAG	CAGCTTCAAT	AATACTGGAG	TTTTTTCTCA	TAGTTTTCCT	TATTCACAA
5041	CCAGAAAAGC	AGAGAAACCC	CCAAGATAAC	CAATTGACCT	ACGTTGTCAAT	AGCCATCTCT
5101	ACAGTGGTGG	CCGCAACCAAT	GGCAAAACGAG	ATGGGTTTCC	TGSAAAAAC	GAAGAAAGAT
5161	CTCGGATTGG	GAAGCATTAC	AACCAGCAA	CCCCAGAGCA	ACATCCCTGCA	CATAGATTA
5221	CGTCCCGCAT	CAGCATGGAC	GCTGTATGCT	GTGGCCACAA	CATTTGTCAAC	ACCAATGTTG
5281	AGACACAGCA	TTGAAAATTC	CTCAGTGAAC	GTGTCCCTAA	CAGCTATTGC	CAACCAAGCC
5341	ACAGTGTTAA	TGGGTCTTGG	GAAAGGATGG	CCATTGTCAA	AGATGGACAT	CGGAGTTCCT
5401	CTTCTCGCCA	TTGGATGCTA	CTCACAAAGTC	AACCCCATAA	CTCTCACAGC	AGCTCTTTT
5461	TTACTGGTAG	CACATTATGC	CATCATAGGG	CCAGGACTCC	AAGCAAAAACC	AACCAGAGAA
5521	GCTCAGAAAA	GAGCAGCAGC	GGGCATCATG	AAAAACCCAA	CTGTCCGTAG	AAIAACAGTG
5581	ATTGACCTAG	ATCCAATACC	CTATGATCCA	AAGTTTGAAA	AGCAGTTGGG	ACAAGTAAAT
5641	CTCTAGTCC	TCTCGGTGAC	TCAAGTGTGG	ATGATGAGGA	CTACATGGGC	TCTGTGTGAG
5701	GCTTTAACCT	TAGCGACCGG	GCCTATCTCC	ACATTGTGGG	AAGSAAAACC	AAGGAGTTT
5761	TGGAACACTA	CCATTGCAAT	GTCAATGCT	AACATTTTAA	GAGGGAGTTA	CTTGGCCGGA
5821	GCTGGACTTC	TCTTTTCCAT	CATGAAGAAC	ACAACCAACA	CGAGAAGGGG	AACTGGCAAC
5881	ATAGGAGAGA	CGCTTGGAGA	GAAATGSAAA	AGCCGATTGA	ACCCATTGGG	GAAAAGTGA
5941	TTCCAGATCT	ACAAGAAAAG	TGGAATCCAG	GAAGTGATA	CAACCTTAGC	AAAAGAAGCC
6001	ATTAAAAGAG	GAGAAAACGA	CCATCACGCT	GTGTCCGGAG	GCTCAGCAAA	ACTGAGATCC
6061	TTCTGCGAGA	GAAATATGGT	CACACCAGAA	GGGAAAAGTAG	TGGACCTCCG	TTGGGGCABA
6121	GGAGGCTGGT	CATACTATTG	TGSGGGACTA	AAGAATGTAA	GAGAAGTCAA	AGGCCTAACA
6181	AAAGGAGGAC	CAGGACATGA	AGAACCCTATC	CCCATGTCAA	CATATGGGTG	GAATCTAATA
6241	CGTCTTCAAA	GTGGAGTTGA	CGTTTTCTTC	ACTCCGCCAG	AAAAGTGTGA	CACATTTGTC
6301	TGTGACATAG	GGGAGTCGTC	ACCAAAATCC	ACGGTAGAAG	CAGGAAGAAC	ACTGAGATTC
6361	CTTAACCTAG	TGGAAAATTC	GTTGAACAAC	AACACCCAAT	TTTGCATAAA	GGTCTCAAC
6421	CCATACATGC	CCTCAGTCAAT	AGAAAAAATG	GAAGCACTAC	AAAGGAAATA	TGAGAAACC
6481	TTAGTGAGGA	ATCCACTTTC	ACGAAAATCC	ACACATGAGA	TGTACTGGGT	ATCCAATGCC
6541	TCCGGGAACA	TAGTGTCAAT	AGTGAACATG	ATTTCAAGGA	TGTTGATCAA	CAGATTCACA
6601	ATGAGACACA	AGAAAGCCAC	TTACGAGCCA	GATGTAGACC	TGGGAAGCGG	AACCCGCAAC
6661	ATCGGAATTG	AAAGTGAGAT	ACCAAACCTA	GACATAATCG	GGAAAAGAAT	AGAAAAATA
6721	AAACAAGAGC	ATGAAACATC	ATGGCACTAT	CACCAACACC	ACCCATACAA	AAAGTGGGCT
6781	TACCATGGCA	GCTATGAAAC	AAAACAAACT	GGATCAGCAT	CATCCATGGT	GAACGGAGTG
6841	GTCAGACTGC	TGACAAAACC	TTGGGACGTC	GTCCCCATGG	TAACACAGAT	GGCAATGACA
6901	GACACGACTC	CATTTGGACA	ACAGCGCGTT	TTTAAAGAAA	AAGTGGACAC	GAGAACCCAA
6961	GAACCGAAAG	AAGGCACAAA	GAAACTAATG	AAAATCACGG	CAGAGTGGCT	TTGGAAAGAA
7021	CTAGGGAAGA	AAAAGACACC	TAGGATGTGC	ACTAGAGAAG	AATTCACAAG	AAAGGTGAGA
7081	AGCAATGCAG	CCTTGGGGGC	CATATTTACT	GATGAGAACA	AGTGGAAAGTC	GGCACGTGAG
7141	GCTGTTGAAG	ATAGTAGGTT	TTGGGAGCTG	GTTGACAAGG	AAAGGAATCT	CCATCTTGAA
7201	GGAAAGTGTG	AAACATGTGT	GTATAACATG	ATGGGAAAAA	GAGAGAAGAA	GCTAGGGGAG
7261	TTCGGCAAGG	CAAAAGGCAG	CAGAGCCATA	TGGTACATGT	GGCTTGAGC	ACGCTTCTTA
7321	GAGTTTGAAG	CCCTAGGATT	CTTGAATGAA	GATCACTGCT	TCTCCAGAGA	GAACTCCTTG
7381	AGTGGAGTGG	AAGGAGAAGG	GCTGCACAAG	CTAGGTTACA	TTTTAAGAGA	CGTGAGCAAG
7441	AAAGAGGG					

Fig. 3

2

Fig. 4

1	QEGAMHTALT	GATEIQMSSG	NLLFTGHLKC	RLRMDKLQLK	GMSYSMCTGK	FKVVKEIAET
61	QHGTIVIRVQ	YEGDGSFCKI	PFEIMDLEKR	HVLGRLITVN	PIVTEKLSIV	NIEAEPFGD
121	SYIIIGVEPG	QLKLNWFKKG	SSIGQMIETT	MRGAKRMAIL	GDTAWDFGSE	EECLTSIGKA
181	LHQVFGAIYG	AAFSGVSWIM	KILIGVIITW	IGMNSRSTSL	SVSLVLVGVV	TLYLGVMVQA
241	DSGCVVSWKN	KELKCGSCIF	ITDNVHTWTE	QYKFQPEPSP	KLASAIQKAH	EEGICGIRSV
301	TRLENLMWKQ	ITPELNHILS	ENEVCLTIMT	GDIKGIQAG	KRSLQFPQTE	LKYSWKTWGG
361	AKMLSTESHN	QTFLIDGPET	AECFNTNRAW	NSLEVEDYGF	GVFTTNIWLK	LREKQDVFCG
421	SKLMSAAIKD	NRAVHADMGY	WIESALNDTW	KIEKASPIEV	KSCHWPKSHT	LWENSVLESL
481	MIIPKNFAGF	VSQHNRYRFGY	HTQTAGPWHL	GKLEMDFFDC	EGTTVVVTEG	CCNRGFSLEK
541	TTASGKLITE	WCCRSCTLFF	LBYRGEDGQW	YGMEIRPLKE	KEENLVNSLV	TACHGGIINF
601	SLGVLGMALF	LEEMLRTRVG	TKHAILLVAV	SFVTLITGNM	SFRDLGRVMV	MVGATMTDDI
661	GMGVTYLALL	AAFKVRPTFA	AGLLLRKLTG	KELMMTTIGI	VLLSQSTIFE	TILELTDALA
721	LGMVMVKMVR	KMEKYQLAVT	IMAILCVINA	VILQNAWKVS	CTILAVVSVS	PLFLTSSQKQ
781	ADWIPLALTI	KGLNPTAIFL	TLSRRTNKKR	SWPLNEAIMA	VGMVSLASD	LLKNDIPMTG
841	PLVAGCLLTV	CYVLTGRSAD	LELERAADVK	WEDQAEISGS	SPILSITISE	DGSMISIKNEE
901	EEQTLTILIR	TGLLVISGLF	PVSIPITAAA	WYLWEVKKQR	AGVLWVVPF	FPVKGABELD
961	GAYRIKQKGI	LGYSQIGAGV	YKEGTFHTMW	HVTRGAVLMH	KCKRIEPEWA	DVKKDLVSCG
1021	GGWKLEGEWK	EGEEVQVLAL	EPGKNTRAVQ	TKFGLFKTNA	GTIGAVELDF	SPGPEGSEPII
1081	DKKGVVGLY	GNGVVTRSGA	YVSAIAQTEK	SIEDNPEIED	DIFKRRKLLI	MDLHPCAGKT
1141	KRYLPAIVRG	AIKRGRLTLI	LAPTRVVAEE	MEEALRGLPI	RYQTPAIRAE	HTPREIVDLM
1201	CHATFTMELL	SPVRVPNYNL	IIMDEAHFTD	IASIAARGYI	STRVEMGEAA	GIFMTATPFG
1261	SRDPFPQSNA	PIMDEEREIP	ERSWSSGHEW	VTDPKGKTIV	FVFSIKAGND	TAACLKKNKG
1321	KVTQLSRKTF	DSEYVKTRTN	DWNFVVTDI	SEMGANFKAE	RVEDPFRCKM	FVILTDGLEH
1381	VILAGFMPVT	HSSAAQRGR	IGRNPKNEND	QYIYMGEPL	NDEDCAHWKE	AKMLLENINT
1441	PEGIIFSMFE	PEREKVDAID	NEYRLRGEAR	KTFVDLMRRG	DLFVWLAYRV	AAEGINYADF
1501	RWCDFGIKNN	QILEENVEVE	IWTKEGERKK	LKPRWLDAMI	YSDPLALKEF	KEFAAGRCEL
1561	TLNLITEMGR	LPTFMTQKAR	DALDNLAVLH	TAEAGGRAYN	HALSELPETL	ETLLLLTLLA
1621	TVTGGIFLFL	MSGRGIGKMT	LGMCCIIIAS	ILLWYAQIQP	HWIAASIIIE	FFLIVELIFE
1681	PEKQRTPDN	QLTYVVIAIL	TVVAATMANE	MGFLEKTKKD	LGLGSIITGQ	PEGNILDIBI
1741	RFASAWTLYA	VATTFVTPML	RHSIENSSVN	VSLTAIANQA	TVLMGLGKGW	PLSKMDIGVF
1801	LLAIGCYSQV	NPITLTAALF	LLVAHYAII	PGLQAKATRE	AGKRAAAGIM	KNPTVGRITV
1861	IDLDPIDYDP	KFEKQLGQVM	LLVLCVTVL	MMRTTWALCE	ALTATGFI	TLWEGNPGRF
1921	WNTTIAVSMA	NIFRGSYLAG	AGLLFSIMKN	TTNTRRGTON	IGETLGEKWK	SRINALGKSE
1981	FQIYKKSQIQ	EVDRTLAKEG	IKRGETDHHA	VSRGSAKLRW	FVERNMVTP	GKVVDLGGG
2041	GGWSYYCGGL	KNVREVKGLT	KGGPGHEEPI	PMSTYGNLV	RLQSGVLVFF	TPPEKCDTLL
2101	CDIGESSPNP	TVEAGRTRLV	LNLVENWLN	NTQFCIKVLN	FYMPSVIEKM	EALGRKYGGA
2161	LVRNPLSRNS	THEMYWVSNA	SGNIVSSVM	ISRMLINRFT	MRHKKATYEP	DVLLGGGARN
2221	IGIESEIPNL	DIIGREIEKI	KQEHETSGHY	DQDHFYKTTA	YHGSYETKGT	GSASSMNGV
2281	VRLLTKPWDV	VPMVTQAMAMT	DTPFGQQRV	FKEKVDTRTQ	EFKEGTTKLM	KITAVWLWKE
2341	LGKKKTPRMC	TREEFTRKVR	SNATLGTIFT	DENKWSARE	AVEDSRFWEL	VDKERNLHLE
2401	GKCETCVYNM	MGKREKKLGE	FGKAKGSRAI	WYMWLGARFL	EFEALGFLNE	GSLVLQRELL
2461	EWSGRRRAAQ	ARLHFKRREQ	ERG			

Fig. 6 ä

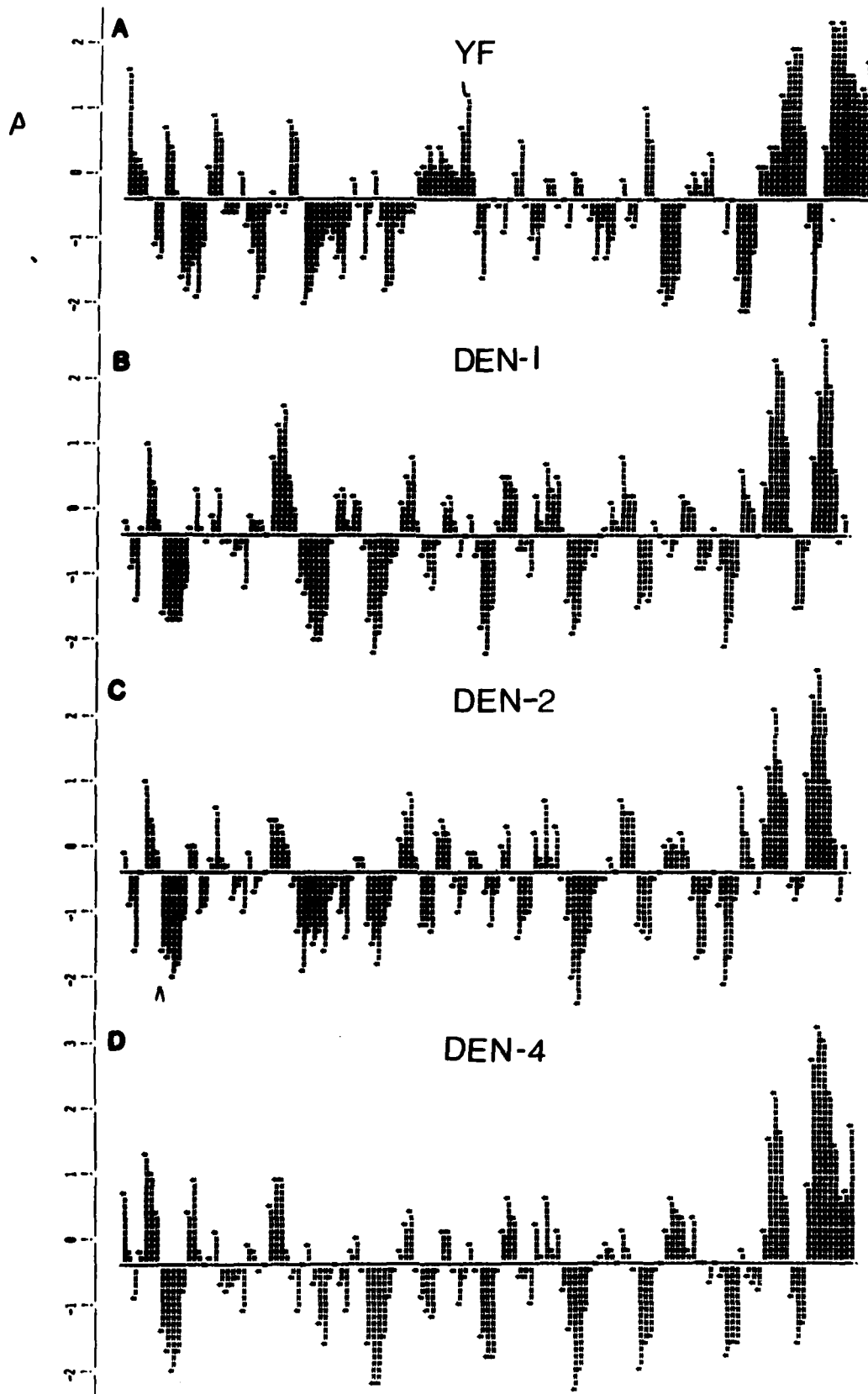
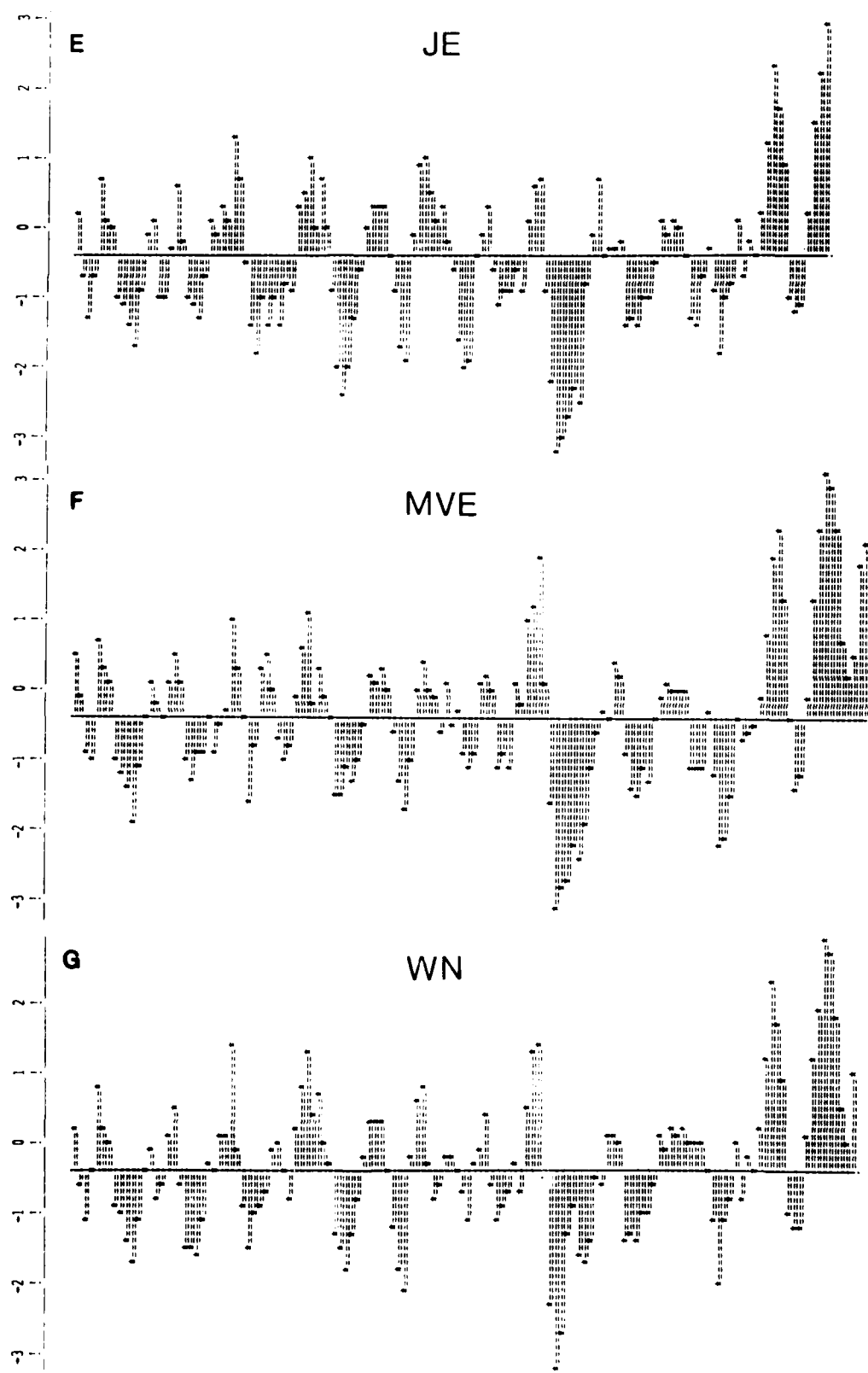


Fig. 6b



1

Fig. 7



602

652

702

752

802

852

902

952

1002

1052

1102

1152

1202

1252

1302

1352

1402

1452

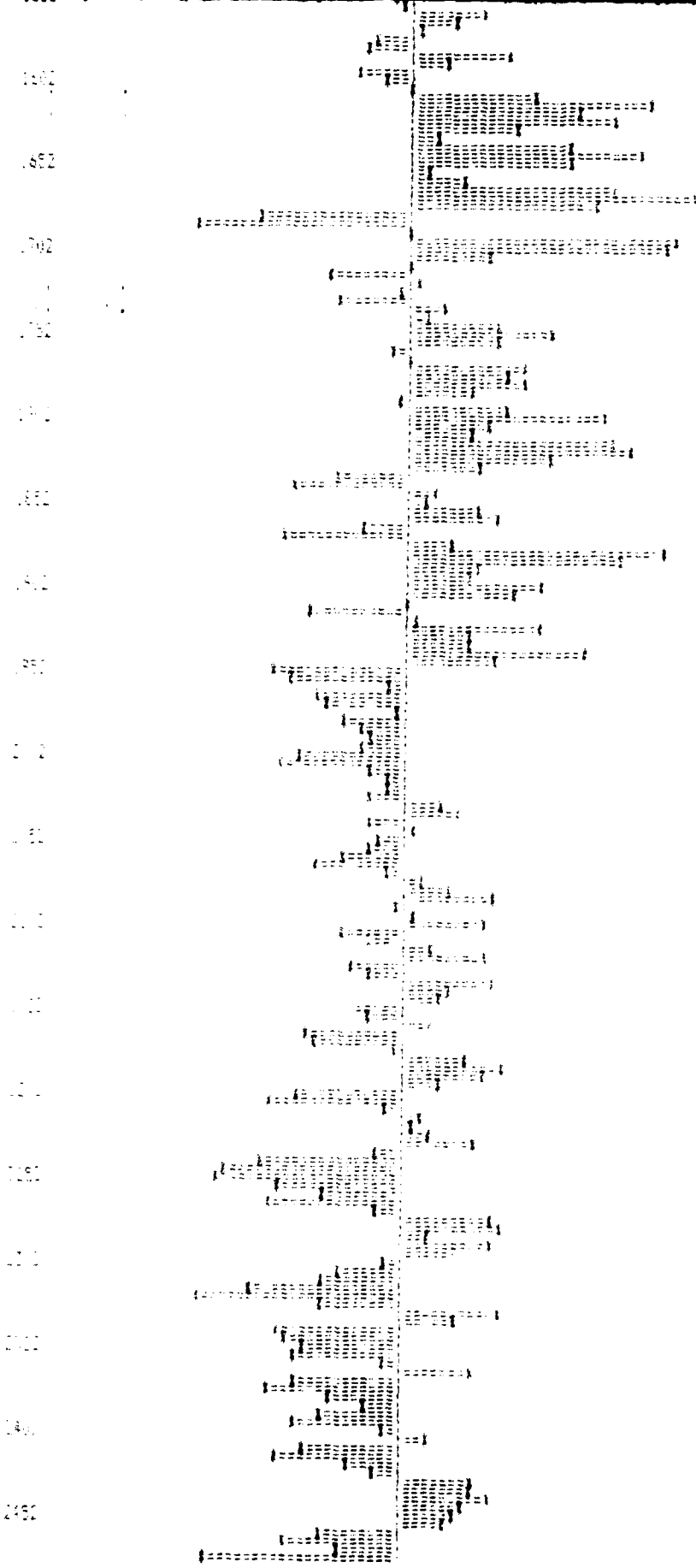
1502

1552

1602

[The following text is a dense, repetitive sequence of characters, likely a data dump or a corrupted document. It consists of numerous lines of alphanumeric characters, some of which are indistinguishable from noise. The text is organized into vertical columns, with a prominent vertical line of characters running down the center. The overall appearance is that of a highly structured but illegible data stream.]

2
1



Plotted from AA 1 to AA 2497
Numbered from AA 1

Fig. 10

```

ns4a
D2AA1 ALKFEKFAAQRKSLTINLITEMRRLPTFTTCHARDALONIAVLHTAEGRAYNHALSELPETLETLILLTLATVTCGIFLFLMSGRGIGKMTIGMCCIIITASILLWYACIIPRINMA
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL ALKSFKDFASGRKSOIG-LVEVLGRMPEHEFMVKTWEALDTMYVVAETAEKGGRAHRMALEELPDALQTVIILALLSVMSLGVFFILMQRKGLGKIGLGGVILGAATFFCWAEEVPGTRIAS
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DCAA1 SIILFEFLIVLLIPEPEKQTRQDNQLTYVVAIALTVVAATMANEMGFLEKTKDGLGSIITQCPESNLDI-----DLRPAASAMTLYAVATTFVPMLRHSIENSVAWSITAIANQA
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL MLLLSLLIMIVLIPESKORSQSTNQLAVFLICVLTLVGAVAAEMGWLDKTKNDIGSILGHRPEAKETTLGVESEFLDLRPAATANSLYAVTAVLTPLLKHLITSDYINTSILTSIN'QA
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
ns4b
D2AA1 TVLMGLGKWPLSKMDIGVPLLAFCYSCVNPPIILTAALFLLVAHYAIIIGPGLQAKATRAAKRAAJIMKNPTVGRITVTDLLPIPYD-PKFEKQLGQVMLLVLCVTQVLMKRTTALJC
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL SALTTLARGFPFVDVGSALLLAVSCWGGVTLTVVTRAAALFCHAYMVPGWQAEAMRQAORRTAAGIMKNVVVDGIVATDVELEERTTPYMQKVGQIILILVMAAVVWVWVNSVTRVR
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
NS5
D2AA1 EALTLAIGPISITLWEGNPGRFWNTTAVGMANTFRGSLAGAGELFSIMKNVTTNRRGTGNI-GETLGEKWKSRNALGLKSEFCIVAKKSGQCEVDRTLAKEGIKRGE-TDHEAVSRGSAR
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL EAGILTTAAAVTLWENGASSVWATTAIGLCHYRRRWLSOLSIMWTLIKNMEKPOLKRGGAKGRTLGEVMKERLNHMTNEEFTRYRKEAITEVDRSNAKHARREGNITGGHRVSRGTAH
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2AA1 LRWFVERNMVTEGKVDLSCCHPQKWSYDGGGLANVREVKGLTKGGPGHEEPIPVMSTYGNLVRLQSGVDVFTTPEPKCDLDCIGSSSNPTVEAGRTLRVNLNVENWLNNTQ-FCI
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL LRWLVERRFLEPVGKVVDLGGCRGGWSYVATQXRVQEVKGYTKGGPGHEEPQLVGSYGNVIVTKSGLDVYRPSASDILLDCIGSSSAAEVEEHRVTVLEWEDWLERGPKRPFCL
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2AA1 KYLNPNPVSVIERKMEALQRKSGADVRNPLSRNSTHEMYVWSNASGNIVSSVNMISRMILINRFTVYRHKKATY-EPDVDLGSOTRNIGIESEIIPNLDIIGKRIEKIKQEHETSKWHDGDRF
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL KYLCPNPKVIEKMETLQRYGASLIENPLSRNSTHEMYVWSHASGNIVHSVMTSQVILGRMEKKTWKGQPFEEVDNLSGSPAVGKPLNSDTSNKKRERLRLKKEYSSTWHDQDHRP
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2AA1 YATWAYHGSYETKQGSASMTVYVPLLTKPVDVVMVTQMAFDTTTFGQQRVKPKVDTRTQEPKGTKKLMAITAVMLWFKYJLKKYTPRMCREEFTKRVRSNATLGTFTDSEWYK
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL YATWNYHGSYEVKPTQGSASMTVYVPLLTKPVDVVMVTQMAFDTTTFGQQRVKPKVDTRTQEPKGTKKYVLAETTNWVWFFLARCWYRMCREEFTIGKWNNAALLGAVFEERQW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DCAA1 KSAREAVEDSRWELVDKERNLMEIKGEGTQNNYVGRKPKKLGEGKAKGSRAWYVWVLAGARFLEFEALGFLNEDHWFSENGLSGVTREGLHKLOYTLRQVSKYKE
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KNVAAL KNAREAVEDPKQEWVWDEBEREALRBEQNTDLYWVMGRKPKPGEFGKAKGSRAIWFVWVLAGARFLEFEALGFLNEDHWFSENGLSGVTREGLHKLOYTLRQVSKYKE
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
-----

```


Fig. 13

```

DEN2 1546 ALKEFKFAAGRSKLTLLMLITEMGRLPFTMQKARDALDNLAVLHTAEAGRAYNHALSELPETLETLLELLLATVTTGGIFLFMEGRGIGKMTLGMCCIIITASILLWYAQIQPHWIAA
YF 2094 SALSEFIKFAEGRGAELVWLSLSELPDFLAKKGEAMDTISVFLHSEEGSRAYRNALSWPEAMTIVMLFILA GLITSGWIFFMSPKGIERSMSMAMGTMACCGYIMFLGGVKPETHISY
DEN2 1666 SIILEFFLIVLLIPEPERQRTPODNQLTYVWIAIITVVAATMANEMGFLEKTKKDLGLSGITTOQPE.SNIL—DIDLRPASAWLTVAVATTFVTPMLRHSIENSSVNVSLTALANQATVL
YF 2214 VMLIFFVLMVVIPPEGQORSIQDNQVAYLIIGILTLVSAVAANELGMLKTKEDLFGKKNLIPSSASPWSWPDIDLKPGAAWTVVGVIVTMLSPLMHHWIKVYGNLSLGGIAQASVZ
DEN2 1784 MGLGKGMPLSKMDIGVPLLAIGCYSQVNPITLTAALFLVAHYAIIIGPLQAKATREAAQRAAAGIMKNPTVGRITVIDLDPIDPYDPKF—EKOLGQVMLLVLCVTVQVLMRMTTVALCEAL
YF 2334 SFMDKGIFFMKNMISVIMLLVSGWNSITVMPLLCGIGCAMLHWSLILPGIKAQSKLAQRVDFHVAENPVVDGNPTVDIEEAPEMPALYKRLALYLLALLSLSAVAMCRTPFSLAEGI
DEN2 1903 TLATGPISTLWFGNPGRFWNTTIAVSMANIFRGSYLAGALLFSIMKNTTIRRGIGNIGETLGEKWKSRNALGKSEFQIVKXSGIQEVCRTLAKEGKRGEDHH—AVSRGSKCRWF
YF 2454 VLASAALGPLIEGNTSLLMNGPMVAVSMTGVWRGNHYAFVGVMMNLWKMKTGRRGSAN—GKTLGEVWKRELNLLDKRQFELLYKRTDIVEVDQDTARRHLAEGKVDVTGVAVSRGTAKLRAF
DEN2 2022 VERNMVTPEGKVVLDGCGRGWSYYCGGLKNVREVKGLTKGGPGHEEPIPMSTYGNLVRLOSGVDVFFTPPEKCDTLLCDIGESSNPFTVEAGRTLRLVNLVENWLNNTQ—FCIKVLN
YF 2572 HERGYVLEGRVIDLCCGRGWCYYAAAQKEVSGYKGFILGRDGHCKPMNVQSLGWNIIIFDKTDIHRLEPVKCDLCCDIGESSSSVTEGERTVRLVDTVEKWLACGGVDNFCVKVLA
DEN2 2141 PMPYSVIEKWEALQRYCGALVRNPLSRNSTHEMYWVSNASGNIVSSVNMISRLINRFTMRHKATYEPDVIDLGSSTRNIGIESEIPNLDIIGKRIEKIQEHETSWHYDQDHPYKQWA
YF 2692 PMPDVLKLELQRRFGGTIVRNPILSRNSTHEMYVSGARSNVITVNTSRLMRRRPTGKVTLEADVILPIGTRSVETDKGPLDKEAIEERVERIKSEYMTSWFYDNDNPRYRW
DEN2 2261 YHGSYETKQTSASSMVNGVVRLLTKPMDVVPVMTQAMTDITPPGQORVFEKVDVTRTOPKKEGTKKIMKITAEMLWMLKELGKKKTPRMCTREEFFTRKVRSNALGALIFTDENKWKKSARE
YF 2812 YCGSYVTKTSGSAAAMVNGVVKILTYPDRIEYVTRMAMTDITPPGQORVFEKVDVTRTOPKKEGTKKIMKITAEMLWMLKELGKKKTPRMCTREEFFTRKVRSNALGALIFTDENKWKKSARE
DEN2 2381 AVEDSRFWEIVDKERNLHLECKGECYVYMMGKREKKGELFCKAKGSRAIWMWMLGARFLEFEALGFLNEDHWF SRENLSLGSVEGEGELHKLGYILRDVSKKE 2483
YF 2932 AVQDPKFWELVDEERKLHQQRGRCYVYMMGKREKKGELFCKAKGSRAIWMWMLGARFLEFEALGFLNEDHWF SRENLSLGSVEGEGELHKLGYILRDVSKKE 3034

```

ns4a region

ns4b region

NS5 region

Cloning Strategy for the DEN2 E/NS1 regions

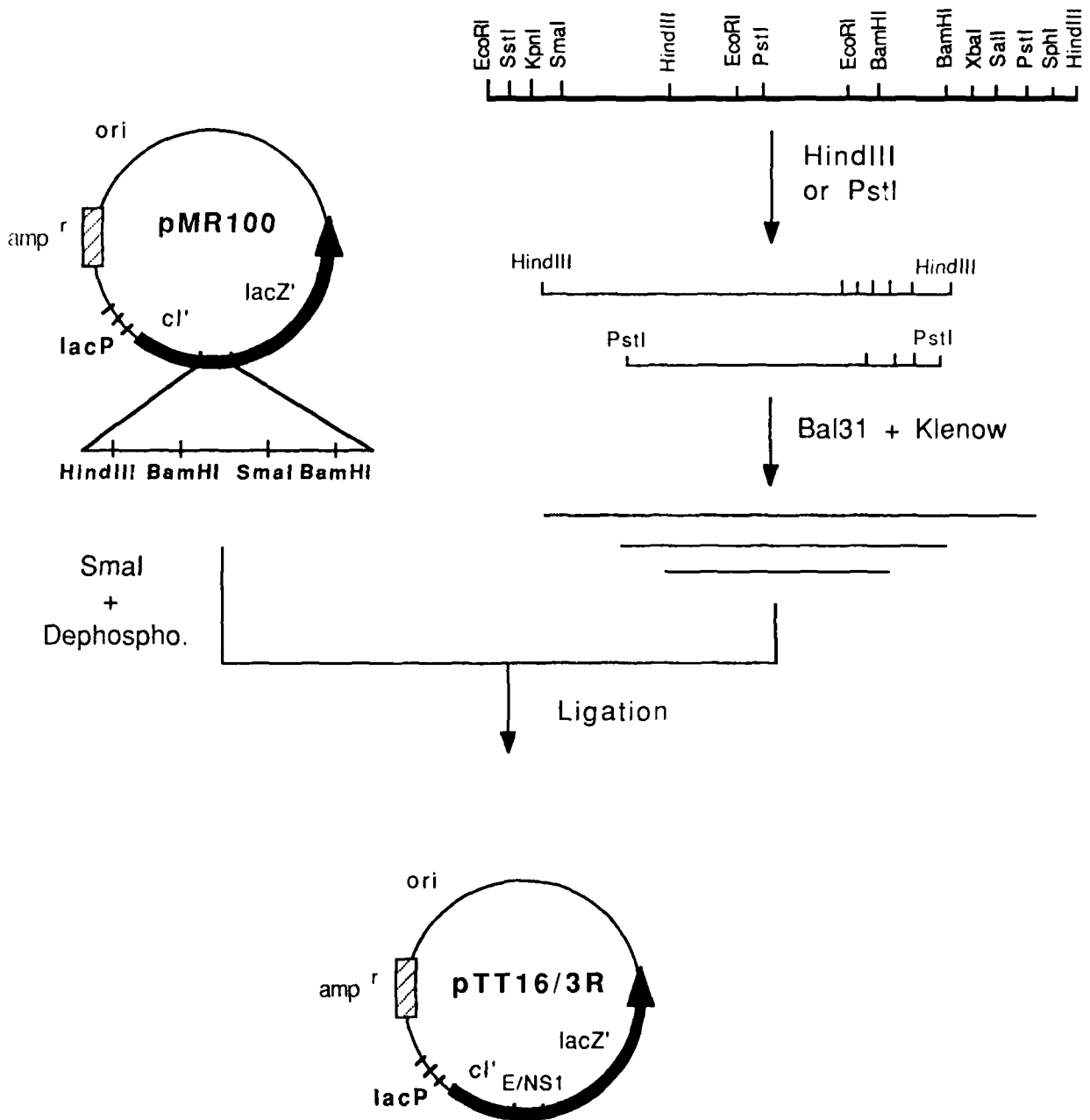


Fig. 14

Fig. 15

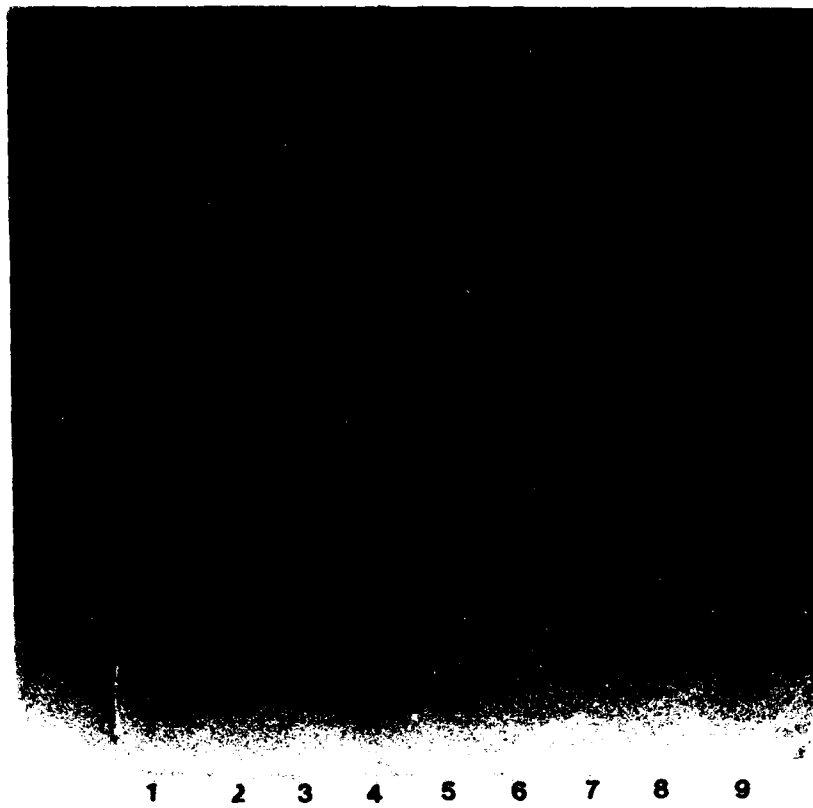


Fig. 16

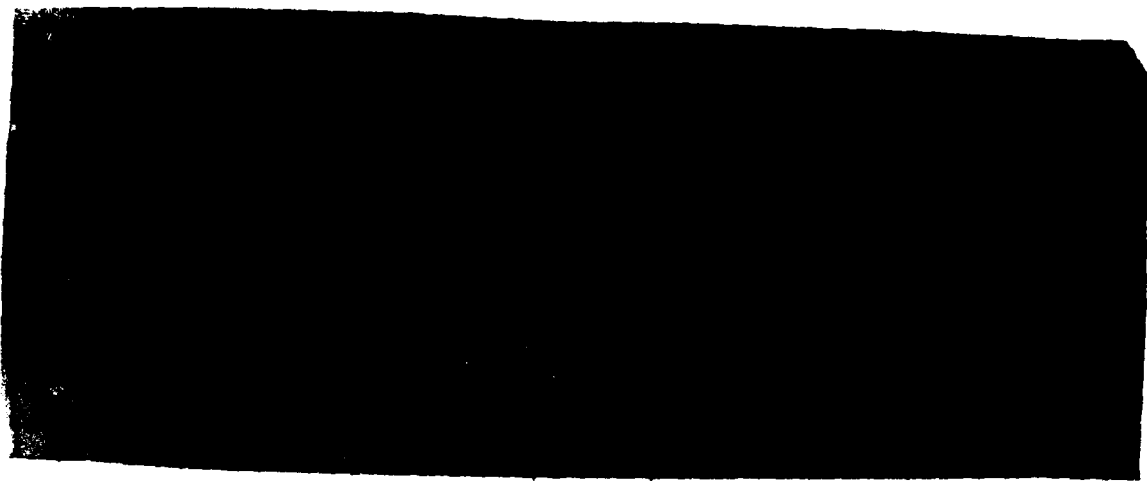


Fig. 17



Fig. 18

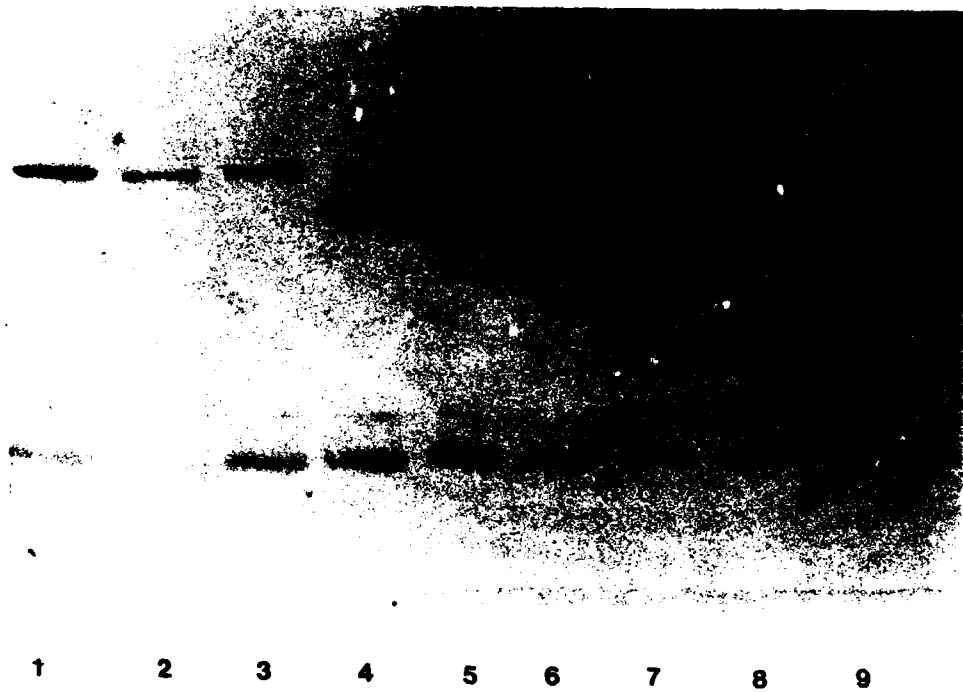


Fig. 19

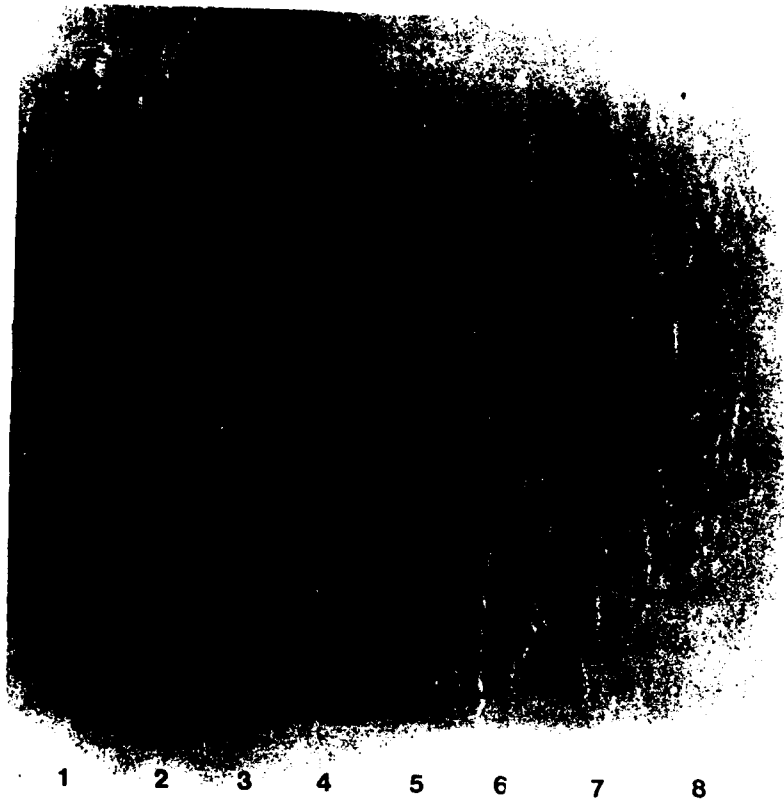


Fig. 20

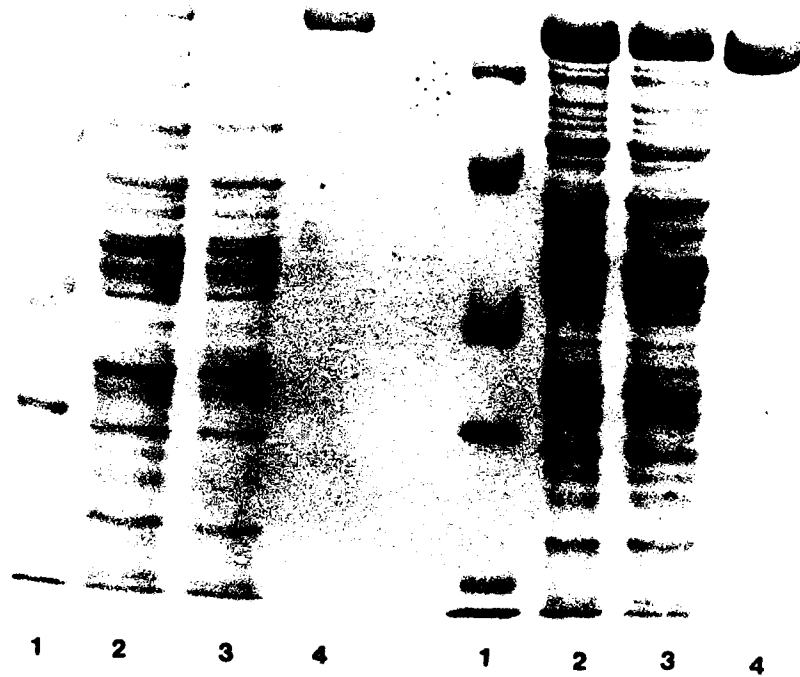


Fig. 21

16

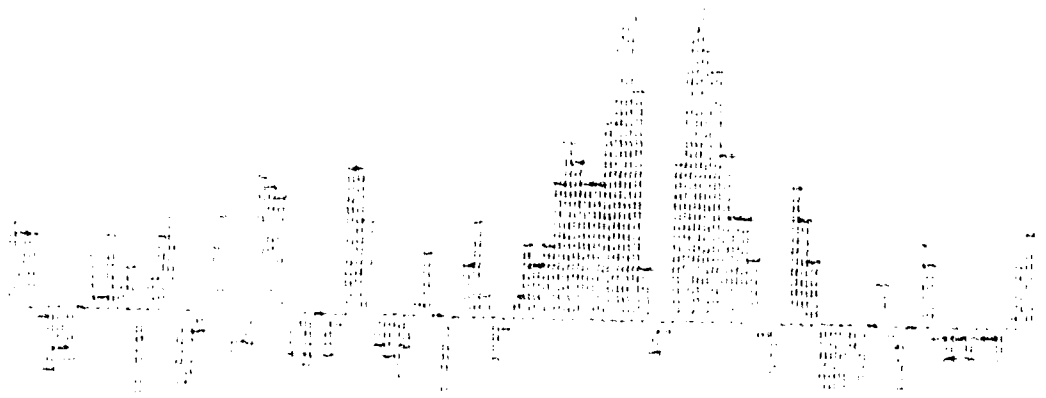
→ E region
DEN2 1 QEGAMHTALTGATEIQMSSGNLLFTGHLKRLRMDKLQKGMSSMCTGKFKVVKETAETQHGTVIRVQYEGDGS PCKIP
DEN2 82 FEIMDLEKRHVLRRLITVNP IVTEK DSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMIETTMRGAK
DEN2 156 RMAILGDTAWDFGSWE ECLTSIGKALHQVFGAIYCAAFSGVSWIMKILIGV IITWIGMNSRSTSLSVSLV LVGV
→ NS1 region
DEN2 270 VTLYLGVMVQADSGCVVSWKNKELKCGSGIFITDNVHTWTEQYKFPESP SKLASAIQKAHEEGICGIRSVTRL
DEN2 304 ENLMWKQITPELNHILSENEVKLTIMTGD IKGIMQAGKRSLOPQPT ELKYSWKTWCKAKMLSTESHN 369

3R

→ E region
DEN2 1 QEGAMHTALTGATEIQMSSGNLLFTGHLKRLRMDKLQKGMSSMCTGKFKVVKETAETQHGTVIRVQYEGDGS PCKIP
DEN2 81 FEIMDLEKRHVLRRLITVNP IVTEK DSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMIETTMRGAKRMAIL
DEN2 161 GDTAWDFGSWE ECLTSIGKALHQVFGAIYCAAFSGVSWIMKILIGV IITWIGMNSRSTSLSVSLV LVGVVTL YLGVMVQ
→ NS1 region
DEN2 240 ADSGCVVSWKNKELKCGSGIFITDNVHTWTEQYKFPESP SKLASAIQKAHEEGICGIRSVTRLENLMWKQITPELNHI
DEN2 319 LSENEVKLTIMTGD IK 334

Fig. 22

3R



16



38

Fig. 23



Fig. 24



1 2 3 4 5 6 7 8 9 10

Fig. 25

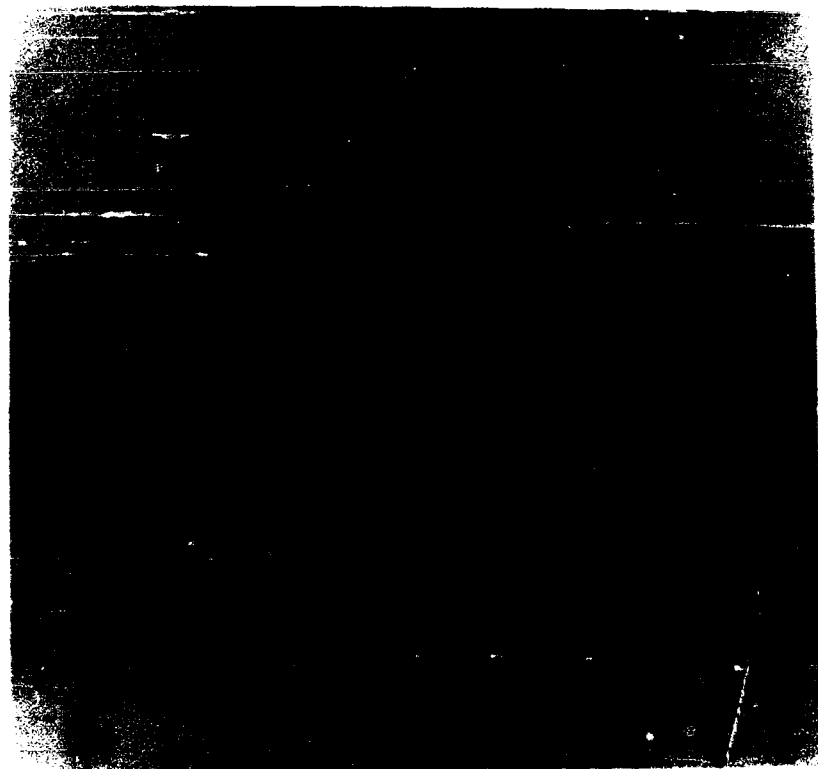


Fig. 26

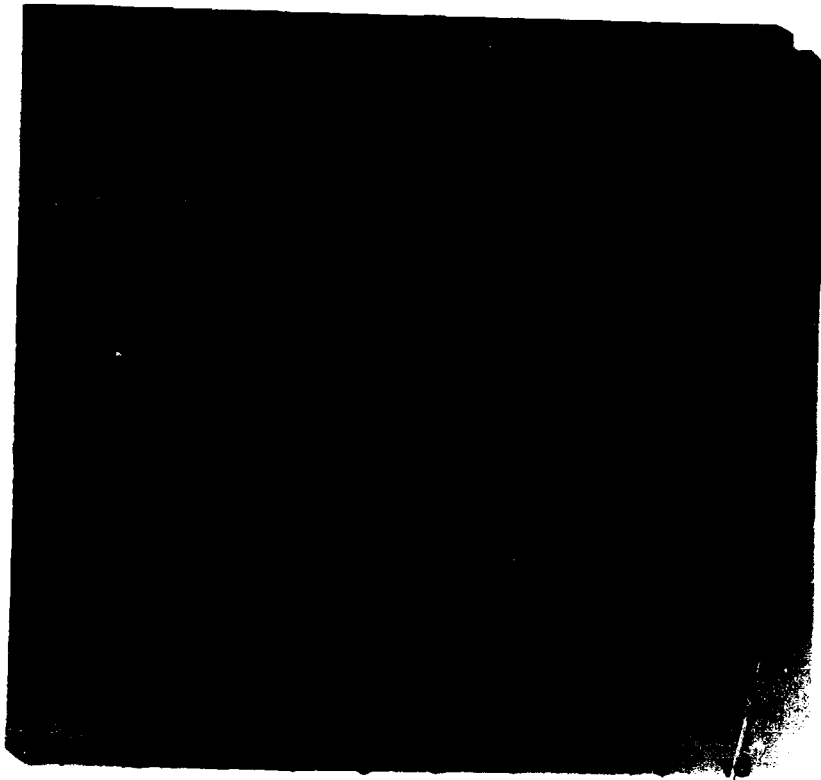


Fig. 27

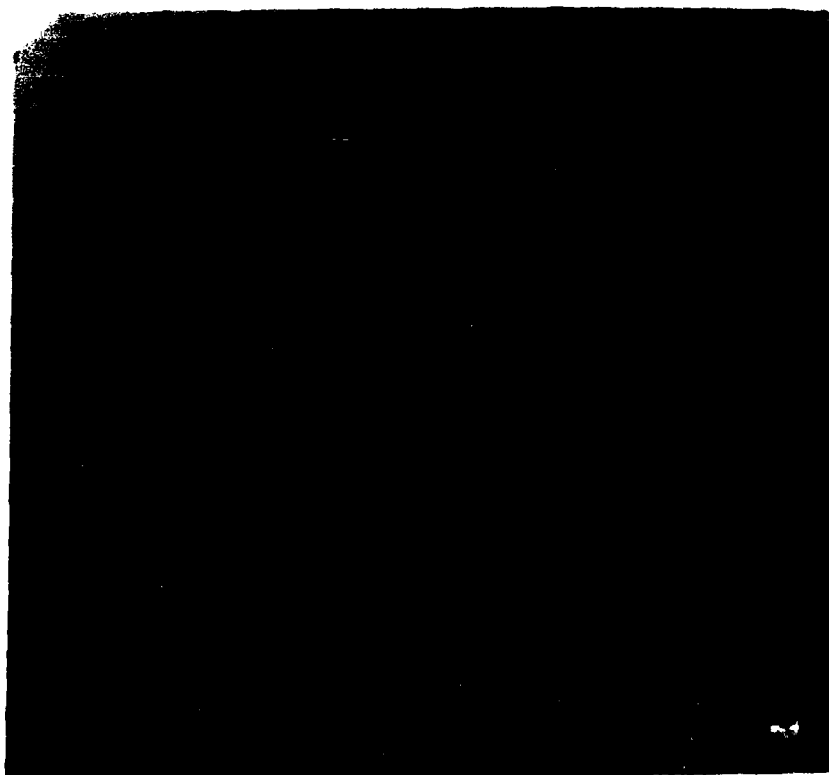


Fig. 28

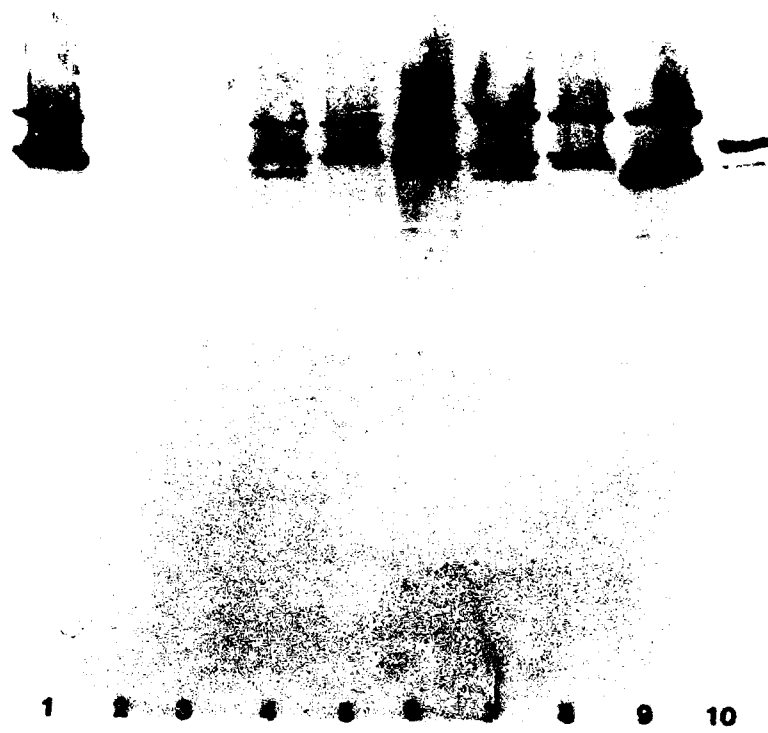


Table I

DEN2 Clones	pRP2	pVY9	pVY1	pVY17		DEN2 Clones	pRP2	pVY9	pVY1	pVY17
B1						A3				
B2						A4				+
B28						A6				
B30						A7				
11						A8				
22						A10				
23						A11				
33						B1				
505			+	+		B2				
522						B3			+	
523		+				B5				
524						B6				
563						B7		+	+	
584		+				B11			+	
614						B12				
620						C2				
621						C3				
635						C4	+			
651						C5				
660						C6				
664						C8				
679		+				C10				
706						C11		+	+	
860	+	+				C12	+			
817		+				D2	+			
S6		+	+			D4				+
S10				+		D6				
S20						D7				
S25						D8				+
S31						D9	+			
S32						D10				+
S43						D12	+			
S44						G3	+			
S46						G5				
S48						G7	+			
S49				+		G8	+			
S53						G9	+			
S60		+	+			G10		+		
S62						G11				
S63		+	+			G12				+
S72						H1				+
S77				+		H2	+			
S78						H3				+
S101						H7		+		
S102						H8				+
S105						H9				
A2						H11				

Table 2

Homology between DEN2:YF and DEN2:WNV

<u>Coding region</u>	<u>Identical a.a.</u>		<u>Total a.a.#</u>	<u>% homology</u>	
	<u>YF</u>	<u>WNV</u>		<u>YF</u>	<u>WNV</u>
NS1	152	213	410	37	52
ns2a	26	25	160	16	16
ns2b	42	31	128	33	24
NS3	298	367	619	48	59
ns4a	102	123	286	36	43
ns4b	38	33	112	34	29
NS5 (partial seq.)	308	326	526	59	62

Literature Cited

- Brandt, W.E., Cardiff, R.D., and Russell, P.K. (1970). Dengue virions and antigen in brain and serum of infected mice. *J. Virol.* 6, 500-506.
- Castle, E., Leidner, U., Nowak, T., and Wengler, G. (1986). Primary structure of the West Nile flavivirus genome region coding for all non-structural proteins. *Virology* 149, 10-26.
- Dalgarno, L., Trent, D.W., Strauss, J.H., and Rice, C.M. (1986). Partial nucleotide sequence of the Murray Valley encephalitis virus genome; comparison of the encoded polypeptides with yellow fever virus structural and non-structural proteins. *J. Mol. Biol.* 187, 309-323.
- Gray, M.R., Colot, H.V., Guarente, L., and Rosbash, M. (1982). Open reading frame cloning: identification, cloning, and expression of open reading frame DNA. *Proc. Natl. Acad. Sci. USA* 79, 6598-6602.
- Henchal, E.A., Henchal, L.S., and Thaisomboonsuk, B.K. (1987). Topological mapping of unique epitopes on the dengue-2 virus NS-1 protein using monoclonal antibodies. *J. of Gen. Virol.* 68, 845-851.
- Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105-132.
- McAda, P.C., Mason, P.W., Schmaljohn, C.S., Dalrymple, J.M., Mason, T.L., and Fournier, M.J. (1987). Partial nucleotide sequence of the Japanese Encephalitis virus genome. *Virology* 156, 348-360.
- Pustell, J., and Kafatos, F.C. (1984). A convenient and adaptable package of computer programs for DNA and proteins sequence management, analysis, and homology determination. *Nuc. Acids Research* 12, 643-655.
- Rice, C.M., Lenches, E.M., Eddy, S.R., Shin, S.J., Sheets, R.L., and Strauss, J.H. (1985). Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* 229, 726-733.
- Schlesinger, J.J., Brandriss, M.W. and Walsh, E.E. (1987). Protection of mice against dengue 2 virus encephalitis by immunization with the dengue 2 virus non-structural glycoprotein NS-1. *J. Gen. Virol.* 68, 853-857.
- Yaegashi, T., Vakharia, V.N., Page, K., Sasaguri, Y., Feighny, R., and Padmanabhan, R. (1986). Partial sequence analysis of cloned dengue virus type 2 genome. *Gene* 46, 257-267.

Personnel Supported from September 15, 1986 - September 14, 1987

<u>Personnel</u>	<u>% Effort in the reporting period</u>	<u>Months</u>	<u>Period</u>
<u>Research Associates</u>			
Dr. Yasuyuki Sasaguri	83	10	September 14, 1986- July 15, 1987
Dr. Kamal Bittar	25	3	September-Dec. 1986
Dr. Gunwar Sripad	87.5	10.5	Nov. 1, 1986- Oct. 31, 1987
Dr. Marutlin Mohan	37.5	4.5	May 1, 1987-present
Dr. Koji Irie	79	9.5	Jan. 1, 1987-present
		<u>Subtotal Months</u>	<u>37.5</u>
<u>Research Assistants</u>			
Dianne Vassmer	100	12	Sept. 15, 1986- present
Karin Page	79	<u>9.5</u>	Nov. 30, 1986- Oct. 7, 1987
		<u>Subtotal Months</u>	<u>21.5</u>
<u>Graduate Student</u>			
Thaweesak Trirawanapong	50	6	Sept. 15, 1986- present
<u>Laboratory Aide</u>			
L.R. Anderson (2.75%)	34	4	
Tina Nguyen (6.25%)			
Cindy Smith (25%)			Sept. 15, 1986- present
<u>Principal Investigator</u>			
Radha K. Padmanabhan	35	4.2	Sept. 15, 1986- present
		<u>Total</u>	<u>73.2</u>